TOLERABILITY, ASSESSMENT, AND PREDICTION OF PSILOCYBIN-INDUCED ALTERED STATES OF CONSCIOUSNESS

Thesis

presented to the Faculty of Arts

of

the University of Zurich

for the degree of Doctor of Philosophy

by

ERICH STUDERUS

of Waldkirch SG

Accepted in the fall semester 2011

on the recommendation of

Prof. Dr. med. Daniel Hell and Prof. Dr. rer. nat. Boris Quednow

Zurich, January 2012

ABSTRACT

Since the early 1990s, hallucinogenic drugs, such as psilocybin, have been increasingly used to investigate the neuronal basis of altered states of consciousness and psychosis. Furthermore, renewed interest has emerged in using these drugs as an adjunct to psychotherapy. Nevertheless, the therapeutic and experimental use of these substances is still controversial due to fears of potential harm. Although the experience of many investigators suggests that potential risks are minimal when these drugs are administered in a carefully monitored clinical or research environment, the subjective tolerability of these drugs under these conditions has not yet been evaluated in large samples. The revival of hallucinogen research during the past 20 years has also greatly increased the need for well-validated instruments assessing the the acute subjective effects of these drugs. Although Adolf Dittrich's questionnaires for the assessment of altered states of consciousness (ASCs) were frequently used for this purpose, especially in Europe, the factorial structure of these questionnaires is not clearly established because previous psychometric investigations have serious methodological limitations. Finally, the effects of hallucinogens are believed to be critically dependent on non-pharmacological variables (e.g., the user's personality, current mood state and environment), but few empirical studies have investigated several of these predictor variables at a time. Thus, little is known about the order of importance of these variables.

To solve these problems, three empirical studies were conducted, all of which were based on pooled data from Prof. Vollenweider's research group at the University Hospital of Psychiatry in Zurich. Vollenweider's group was one on the first to restart human hallucinogen research in the early 1990s and since then has collected an amount of data that is unrivaled in the world.

In the first study, acute, subacute, and long-term subjective effects of psilocybin were investigated by analyzing the pooled data of eight double-blind placebo-controlled experimental studies. The sample included 110 healthy subjects who had received 1-4 oral doses of psilocybin in a dose range of 45-315 µg/kg body weight. It was found that the effects of psilocybin were generally well tolerated. Most subjects described the experience as pleasurable, enriching, and non-threatening. Strong anxiety and/or dysphoria occurred only in the two highest dose conditions in a relatively small proportion of subjects and in all cases resolved by providing emotional support and without pharmacological intervention. Complaints 24 h after drug intake were mild and mostly included headache and fatigue. Furthermore, follow-up interviews conducted 8-16 months after the psilocybin sessions indicated that all of the subjects were healthy and that none of them had experienced any of the most feared negative consequences of hallucinogen exposure, that is, flashbacks, prolonged psychosis, or subsequent drug abuse.

The second study critically examined the psychometric properties of the altered states of consciousness rating scale OAV in a sample of psilocybin (n = 327), ketamine (n = 162), and MDMA (n = 102) induced ASCs. The factorial structure was analyzed by using exploratory factor analysis (EFA), hierarchical item clustering (ICLUST) and

various techniques of structural equation modeling (SEM), including confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM) and multiple indicators and multiple causes (MIMIC) modeling. The results indicated that the originally proposed factorial structure did not fit the data well. An improved model with 11 factors and 42 items provided a much better fit to the data. MIMIC modeling indicated that this factorial structure was sufficiently stable across drugs, settings, questionnaire versions, and sexes. Compared to the original OAV scales, the new OAV scales differentiated better among the three drug groups, were more homogeneous, and had better convergent and discriminant validities.

The aim of the third study was to detect the most important predictors of psilocybin response. The effects of 24 predictor variables were examined in a sample of 409 psilocybin sessions. It was found that drug dose was by far the most important predictor. However, several non-pharmacological variables also played an important role in the effects of psilocybin. Specifically, having a high score in the personality trait "absorption", being in an emotionally excitable and active state immediately before drug intake, having experienced few psychological problems in the past weeks, no previous experience with classical hallucinogens, and moderate THC and alcohol consumption increased the intensity of pleasurable effects and/or visual alterations, whereas settings involving PET measurements, emotional excitability, and low age contributed to the experience of unpleasant and/or anxious reactions.

Taken together, the three studies have demonstrated that psilocybin induced altered states of consciousness in a carefully monitored research environment are generally well tolerated, can be reliably and validly assessed by 11 new subscales of the OAV questionnaire, and are dependent on several non-pharmacological variables.

ZUSAMMENFASSUNG

Seit den frühen neunziger Jahren werden halluzinogene Drogen, wie z.B. Psilocybin, zunehmend zur Erforschung der neuronalen Grundlagen veränderter Bewusstseinszustände und Psychosen eingesetzt. Zudem ist ein erneutes Interesse aufgekommen, diese Substanzen als Hilfsmittel in der Psychotherapie einzusetzen. Dennoch ist der psychotherapeutische und experimentelle Gebrauch dieser Substanzen immer noch umstritten, da unter anderem schädliche Wirkungen befürchtet werden. Obwohl die Erfahrung zahlreicher Forscher gezeigt hat, dass die potentiellen Risiken minimal sind, wenn die Verabreichung in einem sorgfältig überwachten klinischen oder wissenschaftlichen Setting stattfindet, wurde die Verträglichkeit dieser Substanzen unter diesen Bedingungen bisher nie systematisch in einer grösseren Stichprobe untersucht. Die Wiederbelebung der Halluzinogenforschung in den letzten zwanzig Jahren hat auch zu einem erhöhten Bedarf an Instrumenten geführt, die die subjektiven Wirkungen dieser Substanzen reliabel und valide erfassen können. Mit Adolf Dittrichs Fragebögen zur Erfassung veränderter Bewusstseinszustände liegen zwar bereits Instrumente vor, die sich insbesondere in Europa bewährt haben. Die faktorielle Struktur dieser Fragebögen ist aber nicht eindeutig geklärt, da die bisherigen psychometrischen Studien schwerwiegende methodische Mängel aufweisen. Schliesslich besteht bei der Forschung mit

halluzinogenen Substanzen auch das Problem, dass die Effekte wesentlich durch nicht-pharmakologische Faktoren, wie z.B. Persönlichkeitseigenschaften, aktuelle Stimmung oder die Umgebung, moduliert sein können. Welche Einflussfaktoren am wichtigsten sind, ist bisher aber nicht ausreichend geklärt, da frühere Studien oft nur wenige Variablen auf einmal berücksichtigt haben.

Um die oben genannten Probleme zu lösen, wurden drei empirische Studien durchgeführt, welche alle auf den gepoolten Daten von Prof. Vollenweiders Forschungsgruppe an der Psychiatrischen Universitätsklinik Zürich beruhten. Vollenweiders Gruppe hat in den frühen neunziger Jahren als eine der ersten die Forschung mit halluzinogenen Substanzen beim Menschen wiederaufgenommen und seither eine Datenmenge gesammelt, die weltweit einzigartig ist.

In der ersten Studie wurden akute, subakute und langfristige Wirkungen von Psilocybin untersucht, indem acht doppelblinde placebo-kontrollierte Studien gepoolt wurden. Die Stichprobe bestand aus 110 Versuchspersonen, die 1-4 orale Dosen Psilocybin in einer Dosierung von 45-315 µg/kg Körpergewicht verabreicht bekamen. Es zeigt sich, dass die Probanden die Psilocybinwirkung im Allgemeinen gut tolerierten. Die meisten Probanden erlebten die Wirkung als angenehm, bereichernd, und nicht bedrohlich. Starke Angstzustände und/oder negative Gestimmtheit traten nur in den höchsten zwei Dosierungen in einem relativ kleinen Prozentsatz der Leute auf und liessen sich in allen Fällen durch emotionale Unterstützung und ohne medikamentöse Intervention beheben. Beschwerden 24 h nach Drogeneinnahme waren mild und betrafen am ehesten Kopfschmerzen und Müdigkeit. Follow-up Befragungen 8-16 Monate nach der letzten Psilocybin-Sitzung zeigten, dass alle Probanden gesund waren und dass die am meisten gefürchteten negativen Konsequenzen der Verabreichung von halluzinogenen Drogen (Flashbacks, andauernde psychotische Zustände und Induzierung von Drogenmissbrauch) bei niemandem aufgetreten war.

In der zweiten Studie wurden die psychometrischen Eigenschaften des OAV-Fragebogens in einer Stichprobe von Psilocybin- (n = 327), Ketamin- (n = 162), und MDMA- (n = 102) induzierten veränderten Bewusstseinszuständen untersucht. Die faktorielle Struktur wurde mittels explorativer Faktorenanalyse, hierarchischem Item-Clustering, sowie mehreren Techniken der Strukturgleichungsmodellierung, inkl. konfirmatorischer Faktorenanalyse, explorativer Strukturgleichungsmodellierung und multiple indicators and multiple causes (MIMIC)-Modellierung, analysiert. Es zeigte sich, dass die ursprünglich angenommene Faktorenstruktur nicht gut zu den Daten passte. Ein neu entwickeltes Modell mit 11 Faktoren und 42 Items liess sich wesentlich besser mit den Daten in Einklang bringen. Eine MIMIC-Modellierung bestätigte, dass die neu gefundene Struktur eine genügend hohe Stabilität über die drei Substanzgruppen, Umgebung der Drogeneinnahme, Geschlechter und Fragebogenversionen hinweg aufwies. Im Vergleich zu den ursprünglichen OAV-Skalen, differenzierten die neuen OAV-Skalen besser zwischen den drei Substanzgruppen, waren homogener und hatten eine bessere konvergente und diskriminante Validität.

Ziel der dritten Studie war es, die wichtigsten Prädiktoren Psilocybin-induzierter veränderter Bewusstseinszustände zu eruieren. Hierzu wurden die Effekte von 24 Prädiktorvariablen in einer Stichprobe von 409 Psilocybin-Sitzungen untersucht. Es stellt sich heraus, dass die Drogendosierung bei weitem der wichtigste Prädiktor war. Mehrere nicht-pharmakologische Variablen spielten aber ebenfalls eine wichtige Rolle.

Namentlich erhöhten eine hohe Ausprägung der Persönlichkeitseigenschaft "Absorption", hohe emotionale Erregbarkeit und Aktivität unmittelbar vor der Substanzeinnahme, eine geringes Erleben von psychischen Problemen in den letzten Wochen, keine Vorerfahrung mit halluzinogenen Substanzen und ein moderater Konsum von Alkohol die Intensität von angenehmen Wirkungen und/oder visuellen Veränderungen, während die Drogeneinnahme in Experimenten mit PET-Messungen, eine hohe emotionale Erregbarkeit und ein niedriges Alter die Erfahrung von unangenehmen Wirkungen und/oder angstvolle Reaktionen förderten.

Zusammenfassend haben die drei Studien gezeigt, dass Psilocybin-induzierte veränderte Bewusstseinszustände in einer sorgfältig überwachten Forschungsumgebung im Allgemeinen gut toleriert werden, reliabel und valide mittels 11 neuer Subskalen des OAV-Fragebogens gemessen werden können, sowie von mehreren nicht-pharmakologischen Faktoren abhängen.

PUBLICATIONS INCLUDED IN THIS THESIS

CHAPTER 3

Studerus, E., Kometer, M., Hasler, F., & Vollenweider, F. X. (2011). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: A pooled analysis of experimental studies. *Journal of Psychopharmacology*, *25*(11), 1434–1452. doi:10.1177/0269881110382466

CHAPTER 4

Studerus, E., Gamma, A., & Vollenweider, F. X. (2010). Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS ONE*, *5*(8), e12412. doi:10.1371/journal.pone.0012412

CHAPTER 5

Studerus, E., Gamma, A., Kometer, M., & Vollenweider, F. X. (2012). Prediction of psilocybin response in healthy volunteers. *PLoS ONE*, *7*(2), e30800. doi:10.1371/journal.pone.0030800

Himmel und Hölle sind im Menschen.
Und es ist so, dass man mit diesem Stoff nun Einblick
bekommt in die eigene Hölle oder den eigenen Himmel.

— Albert Hofmann (2003)

ACKNOWLEDGMENTS

First of all, I would like to thank my PhD thesis supervisor Prof. Dr. med. Franz X. Vollenweider. Franz had already supervised my master thesis and encouraged me to pursue a PhD in his research group. Without his encouragement and trust in me I would probably not have followed up a scientific career, but now I'm very glad I did. I'm also deeply grateful that he gave me the opportunity to analyze and publish his pooled data of more than 20 years of human research with hallucinogenic drugs. Having the largest data set of its kind in the world at my disposal was a huge honor and strongly sparked my interest not only in altered states of consciousness and hallucinogenic drugs but also in methodology because it allowed me to delve into rather advanced statistical analyses that can only be performed with larger data sets. In fact, my interest in statistics has grown so much that it has now become my specialty. Franz – open-minded as he is – has always supported my ideas and allowed me to spend a large proportion of my work time learning new statistical methods and programming languages. I can not imagine that I would have learned the same amount in another research group.

I deeply appreciate the resources that Franz has provided me for so many years with his research group, most notably, the open-minded and inspiring atmosphere that he has created and all the fascinating people that he has brought together, many of which have become good friends over the years. I've never come to a place before where I've met so many intelligent, open-minded, creative, unorthodox, and critically thinking people, and I will never forget the philosophical, scientific, and political discussions that we regularly had at lunch and that deeply inspired my work. My gratitude especially goes to Alex Gamma, Marco A. Benz, Michael Kometer, and Felix Hasler for their great support and the many fruitful discussions that we had. I also would like to thank all other members of Franz Vollenweider's research group who have conducted experiments with hallucinogenic drugs and thereby collected data for my research.

I'm also very grateful to all my friends, especially Steve Ebright, Dave Putnam, Reto Amstad, and Remo Prinz, for their enduring friendship and support with motivational ups and downs during this project. Finally, I would like to thank my family and especially my parents, who always stood behind me even though my PhD project took much longer than initially expected.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

5-HT	5-hydroxy-tryptophan
5D-ASC	Five Dimensions of Altered States of Consciousness
AIC	Akaike's information criterion
ANOVA	analysis of variance
APZ	Abnorme Psychische Zustände

ASC	altered state of consciousness
AUA	Auditory Alterations
AUPI	Augsburg Personality Inventory
BETA	Bewusstseinstrübung und Akustische Halluzinationen
BMI	body mass index
BPD	balanced-placebo design
BPRS	Brief Psychiatric Rating Scale
CFA	confirmatory factor analysis
CFI	comparative fit index
DAE	Derealisation of self and surroundings, Anxious-depressive state, and Euphoric-stimulated state
DED	Dread of Ego Dissolution
D-F	differential functioning
DIF	differential item functioning
DMT	*N*,*N*-dimethyltryptamine
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders - fourth edition
EDN	Experience of Deviation from Normal state
EEG	electroencephalography
EFA	exploratory factor analysis
ESEM	exploratory structural equation modeling
EWL	Eigenschaftswörterliste
EWL-60-S	Eigenschaftswörterliste 60-S
EWL-K	Eigenschaftswörterliste - Kurzform
FFM	Five-Factor Model
FMI	fraction of missing information
FPI	Freiburg Personality Inventory
G-ASC	Global Altered States of Consciouness
GSI	Global Severity Index
HPI-81	Hallucination Prediction Inventory

HPPD	Hallucinogen Persisting Perception Disorder
HRS	Hallucinogen Rating Scale
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
ICLUST	hierarchical item clustering
ISASC	International Study on Altered States of Consciousness
LC	List of Complaints
LR	likelihood ratio
LSD	D-lysergic acid diethylamide
M	mean
MAP	minimum average partial
MAR	missing at random
MCAR	missing completely at random
MCMC	Markov Chain Monte Carlo
MDA	3,4-methylenedioxy-amphetamine
MDE	3,4-methylenedioxy-N-ethylamphetamine
MDMA	3,4-methylenedioxy-methamphetamine
MG-CFA	multiple group confirmatory factor analysis
mGluR2	metabotropic glutamate receptor subtype 2
MI	multiple imputation
MICE	Multivariate Imputation by Chained Equations
MIMIC	multiple indicators and multiple causes
MLR	Robust Maximum Likelihood
NEO-PI-R	Neuroticm, Extraversion, and Openness Personality Inventory Revised
OAV	Ozenaische Selbstentgrenzung, Angstvolle Ichauflösung und Visionäre Umstrukturierung
OBN	Oceanic Boundlessness
PA-PCA	parallel analysis based on principal component analysis
PA-PFA	parallel analysis based on principal factor analysis
PANAS	Positive and Negative Affect Schedule

PASI	Passive Spontaneous Imagination
PCA	principal component analysis
PET	positron emission tomography
POMS	Profile of Mood States
RMSEA	root mean square error of approximation
RIV	relative increase in variance due to missingness
SCL-90-R	Symptom Check-List-90-Revised
SD	standard deviation
SE	standard error
SEM	structural equation modeling
SPSS	Statistical Package for the Social Sciences
SRMR	standardized root mean square residual
STAI	State-Trait-Anxiety Inventory
STAI-S	State-Trait-Anxiety Inventory - State version
TAS	Tellegen Absorption Scale
THC	Δ^9-tetrahydrocannabinol
TLI	Tucker-Lewis index
ULS	unweighted least squares
VAS	visual analogue scale
VIF	variance inflation factor
VIR	Vigilance Reduction
VRS	Visionary Restructuralization
VSS	very simple structure
WLS	weighted least squares
WLSMV	weighted least squares mean and variance adjusted
WRMR	weighted root mean square residual
ZKPQ	Zuckerman-Kuhlman Personality Questionnaire

Part I

INTRODUCTION

1 GENERAL INTRODUCTION

1.1 A BRIEF HISTORY OF PSILOCYBIN RESEARCH

The indolealkylamine psilocybin (4-phosphoryloxy-*N*,*N*-dimethyltryptamine) is the main psychoactive principle of a group of hallucinogenic fungi, commonly known as magic mushrooms. Psilocybin containing mushrooms naturally occur throughout the world and have been used by indigenous people, particularly of Mesoamerica, but possibly also of Papua New Guinea and other regions, in shamanistic rituals and religious ceremonies for centuries, if not millennia (Stamets, 1996; Guzmán, Allen, & Gartz, 2000). Scientific research with psilocybin began with the rediscovery of still active psilocybin mushroom cults by Wasson 1955 (Wasson, 1958), first successful cultivation of psilocybin mushrooms by Heim, and isolation and synthetic reproduction of their pure psychoactive compound by Hoffman (Hofmann, Heim, Brack, & Kobel, 1958). Early human trials in the late 1950s and early 1960s soon discovered that psilocybin exerts similar psychopharmacological actions as other classical serotonergic hallucinogens, such as the semi-synthetic ergoline D-lysergic acid diethylamide (LSD) and the naturally occurring phenethylamine mescaline (e.g., Isbell, 1959; Malitz, Esecover, Wilkens, & Hoch, 1960; Hollister & Hartman, 1962; Wolbach, Miner, & Isbel, 1962).

When Psilocybin was first synthesized, LSD had already been marketed by Sandoz under the name Delysid® for almost a decade and used internationally as a research tool to investigate the neurobiological basis of altered states of consciousness (ASCs) and psychosis and as an adjunct to psychotherapy. However, because psilocybin had a shorter duration of action and seemed to produce less vegetative side-effects, affective disturbances, anxiety and panic reactions (Nieto, 1962; Heimann, 1962; A. E. David & J. M. David, 1961; Clark, 1968; Leuner, 1968; Passie, 1995), it was soon considered by many investigators as a valuable substitute for the earlier discovered LSD. During a short, but fruitful, period from the late 1950s to the late 1960s, at least three dozen research groups in at least two dozen countries have conducted human studies on the effects of psilocybin. Psilocybin was readily accepted in the research community not only due to its favorable action profile, but also because it was marketed by Sandoz under the name Indocybin® for experimental and psychotherapeutic purposes in the early 1960s (Passie, 1995; Passie, Seifert, Schneider, & Emrich, 2002).

However, scientific interest in human experiments with psilocybin and other hallucinogens rapidly declined during the 1960s because an increasing number of people started to believe that many hallucinogenic drug experiments were conducted in an irresponsible manner and that a large amount of the early research was flawed by the lack of controls and inadequate follow-up (Grob, 1994; Hobbs, 2007). Additionally, hallucinogens, particularly LSD, were increasingly used and abused in non-medical settings and associated with the counter culture movement of the 1960s, which claimed to draw inspiration from their use and which was perceived as a threat to society because it promoted social disobedience and anti-authoritarian attitudes (Hobbs, 2007; Wark &

Galliher, 2010). The increased uncontrolled use also led to more adverse drug reactions, such as flashbacks, self-destructive psychotic behavior, and persisting psychosis, which in turn became increasingly the object of sensationalistic media coverage (Grob, 1994; Mangini, 1998; Johnson, Richards, & Griffiths, 2008). Consequently, the public opinion regarding the risks and benefits of hallucinogens dramatically shifted, and by about 1970, hallucinogens were placed into the most restrictive categories of drug prohibition laws in most countries. This meant that these substances were legally defined as having a high potential for abuse, a lack of demonstrated safety, and no accepted medical use. Accordingly, human research with hallucinogens became so difficult and unattractive that it virtually ceased by the early 1970s.

After a near worldwide moratorium of human hallucinogen research for two decades, a growing number of investigators have again started to conduct studies involving hallucinogen administration to human subjects (Langlitz, 2007; Vollenweider & Kometer, 2010). Many of these studies used psilocybin to investigate the neural basis of psychotic symptom formation including ego-disorders and hallucinations (Vollenweider, 1992; Vollenweider, Leenders, Scharfetter, P. Maguire, et al., 1997; Vollenweider, Vollenweider-Scherpenhuyzen, Bäbler, Vogel, & Hell, 1998; Gouzoulis-Mayfrank, Schreckenberger, et al., 1999; Gouzoulis-Mayfrank, Thelen, et al., 1999; Vollenweider & Geyer, 2001; Hasler, Grimberg, Benz, Huber, & Vollenweider, 2004) or to explore the effect of psilocybin on cognitive, visual, and emotional processes (Spitzer, Thimm, et al., 1996; Carter, Pettigrew, Burr, et al., 2004; Carter, Burr, et al., 2005; Carter, Pettigrew, Hasler, et al., 2005; Carter, Hasler, et al., 2007; Kometer, Cahn, Andel, Carter, & Vollenweider, 2011; Schmidt, Csomor, Bachmann, Kometer, & Vollenweider, 2010; Carhart-Harris, Leech, et al., 2012), including mismatch-negativity (Umbricht, Koller, Vollenweider, & Schmid, 2002; Umbricht, Vollenweider, et al., 2003; Heekeren et al., 2008; Schmidt, Bachmann, et al., 2011), sensory gating (Gouzoulis-Mayfrank, Heekeren, Thelen, et al., 1998; Vollenweider, Csomor, Knappe, Geyer, & Quednow, 2007; Quednow, Kometer, Geyer, & Vollenweider, 2012), and time perception (Wackermann, Wittmann, Hasler, & Vollenweider, 2008; Wittmann et al., 2007). Several studies also focused on the pharmacokinetics (Hasler, Bourquin, Brenneisen, Bär, & Vollenweider, 1997), metabolism (Hasler, Bourquin, Brenneisen, & Vollenweider, 2002), dose-response-relationship (Hasler, Grimberg, et al., 2004), tolerability (Carhart-Harris, Williams, et al., 2011), side-effects (Johnson, Andrew Sewell, & Griffiths, 2011) and receptor profile of psilocybin (Ametamey et al., 1998; Vollenweider, Vollenweider-Scherpenhuyzen, et al., 1998; Vollenweider, Vontobel, Hell, & Leenders, 1999; Hasler, Quednow, et al., 2009; Quednow, Treyer, et al., 2011; Quednow, Kometer, et al., 2012).

More recently, some researchers have begun to reevaluate the therapeutic potential of psilocybin (Sessa, 2008; Vollenweider & Kometer, 2010). Conditions that are currently explored and that have responded positively to psilocybin treatment in recent studies include obsessive-compulsive disorder (Moreno, Wiegand, Taitano, & Delgado, 2006), anxiety in patients with advanced-stage cancer (Grob et al., 2011), and cluster headache (Sewell, Halpern, & Pope, 2006). Other researchers have started to evaluate the potential of psilocybin to induce mystical-type experiences and long-lasting positive changes in attitudes, mood, and behavior in healthy volunteers (Griffiths, Richards, McCann, & Jesse, 2006; Griffiths, Richards, Johnson, McCann, & Jesse, 2008; Griffiths, Johnson, et al., 2011; Maclean, Johnson, & Griffiths, 2011). They have found that 72% of subjects had a

"complete" mystical experience, when psilocybin was administered in a very high dose (429 μg/kg) under supportive conditions (Griffiths, Johnson, et al., 2011). Furthermore, at the 14 month follow up, experiences were still rated as having substantial personal and spiritual significance and attributed to positive changes in attitudes, mood, and behavior (Griffiths, Richards, Johnson, et al., 2008; Griffiths, Johnson, et al., 2011).

The research group of Franz X. Vollenweider at the University Hospital of Psychiatry in Zurich has been at the forefront in the revival of human research with hallucinogens. Since the 1990s, this group has conducted 23 psilocybin studies (including pilot studies) using methodology that has not been available in the early phase of hallucinogen research, including modern neuroimaging and electrophysiological protocols. In total, the group has administered psilocybin to more than 260 healthy volunteers in more than 410 individual sessions. Additionally, Vollenweider's group has conducted a relatively large number of studies on the effects of ketamine and MDMA. The amount of data that this group has collected during the past 20 years, particularly with regard to psilocybin, is unrivaled in the world. Fortunately, most of Vollenweider's studies have used similar study designs and rating scales. Thus, they could be easily pooled and served as an ideal basis for the three empirical studies of the present thesis.

1.2 ACUTE PSYCHOLOGICAL EFFECTS

Since psilocybin is usually termed a hallucinogen, one might think that the primary action of psilocybin is to induce hallucinations. However, the effects of psilocybin go far beyond visual alterations; they include altered perception of reality and self, intensification of mood, distorted sense of time and space, activation of vivid memories, enhanced profundity and meaningfulness, and a ubiquitous sense of novelty (Hasler, Grimberg, et al., 2004; Nichols & Chemel, 2006). Furthermore, the visual "hallucinations" induced by psilocybin – at least at moderate doses – are rarely true hallucinations in the sense that they can not be distinguished from real perceptions. Because the deceptive character of the hallucinations is usually noticed, they are more accurately described as non-psychotic or pseudo-hallucinations, although the latter term is problematic because it can be clinically ambiguous (cf. van der Zwaard & Polak, 2001). Thus, many authors have argued that the term "hallucinogen" is actually a misnomer (e.g., Cohen & Ditman, 1963) and have – depending on their specific orientation – introduced alternative descriptions for psilocybin-type drugs (Strassman, 1984). For instance, those who have been struck by the usefulness of these drugs to model aspects of psychosis have introduced the terms psychotomimetics (mimicking psychosis), psychotogens (psychosis inducing), and psychodysleptic (mind disrupting). Others who have used these drugs primarily for psychotherapy and personal growth have called them psycholytic (psyche-loosening), entheogenic (generating the god within), and psychedelic (mind manifesting).

It is clear from the above descriptions that psilocybin and related drugs can induce a variety of very profound alterations of consciousness. These alterations are often highly enjoyable and at high doses can even lead to mystical or so called peak experiences (Pahnke, 1969; Nichols & Chemel, 2006; Griffiths, Richards, McCann, & Jesse, 2006). However, higher doses can also induce terrifying experiences, such as paranoid ideation and cognitive fragmentation (e.g., Malitz, Esecover, et al., 1960; Griffiths, Johnson, et al., 2011). Accordingly, the British writer Aldous Huxley (Huxley, 1959) reported

that these drugs could take one to heaven or hell. Because such powerful alterations of consciousness rarely occur in everyday life, people often have great difficulties to describe them (Nichols & Chemel, 2006). They are also often highly personally meaningful. For instance, in a recent study of Griffiths, Johnson, et al. (2011), in which subjects received five different doses of psilocybin, it was found that 38.9% of subjects rated at least one of the two high-dose psilocybin sessions as as the single most meaningful experience of their life and 77.9% as among the five most meaningful experiences of their life at the 14 month follow-up.

Although several studies have previously investigated acute psychological effects of psilocybin (e.g., Gouzoulis-Mayfrank, Thelen, et al., 1999; Hasler, Grimberg, et al., 2004), there were still many unanswered questions regarding the dose-response-relationships, the cumulative distributions, and time-courses of subjective psilocybin effects. Furthermore, previous studies were often low in sample size. Thus, one aim of this thesis was to investigate acute psychological effects in a larger sample. To this end, psychometric data of eight recent psilocybin studies were pooled and analyzed with a special focus on dose-response-relationships, cumulative distributions, and time-courses of psilocybin effects. The results were published as part of a study that also investigated subacute and long-term psilocybin effects (Studerus, Kometer, et al., 2011). It is included in this thesis as Chapter 3.

1.3 TOLERABILITY OF PSILOCYBIN ADMINISTRATION

Classical hallucinogens in general, and psilocybin containing mushrooms in particular, are considered to have a lower harm potential than most other abused drugs (Nutt, King, Saulsbury, & Blakemore, 2007; Carhart-Harris & Nutt, 2010; van Amsterdam, Opperhuizen, & van den Brink, 2011). For instance, they neither cause any damage to human body organs at regular doses nor engender drug dependence or addiction (Nichols, 2004). Nevertheless, their powerful effects on consciousness can pose serious risks, particularly when higher doses are used recreationally in unsupervised settings. Probably the biggest danger is that acute psychotic reactions might occur, which can lead to dangerous and self-destructive behavior during the time of drug action. Other potential risks of hallucinogen exposure are precipitation or exacerbation of enduring psychiatric conditions, long-lasting perceptual disturbances, commonly known as flashbacks, and development of an abusive pattern of hallucinogen use (El-Mallakh, Halpern, & Abraham, 2008; Hermle, Kovar, Hewer, & Ruchsow, 2008; Johnson, Richards, & Griffiths, 2008).

However, it makes a big difference whether these drugs are used recreationally or administered in controlled research settings where subjects are carefully screened and monitored and judicious doses of pharmaceutical quality drugs are given (Frecska & Luna, 2006). Complications were remarkably rare, when hallucinogenic drugs were used extensively in controlled research settings during the 1950s and 1960s (Cohen, 1960; Strassman, 1984). For instance, flashbacks were virtually never reported in therapeutic or research settings (Halpern & Pope, 2003)

Recently, Johnson, Richards, and Griffiths (2008) have published safety guidelines for human hallucinogen research. The safeguards they propose include the exclusion of volunteers with a personal or family history of psychotic disorders or other severe

psychiatric disorders, establishing trust and rapport between session monitors and volunteers before the session, careful volunteer preparation, a safe physical session environment, and interpersonal support from at least two study monitors during the session. Modern human hallucinogen research has, independently of Johnson's publication, largely adopted these advices, and there is currently no indication that mistakes of early research (i.e., the irresponsible use of some investigators) would be repeated.

Nevertheless, safety and tolerability concerns remain major issues of modern human hallucinogen research. As has been pointed out by Frecska and Luna (2006), many health care providers and legislation makers continue to judge the safety of hallucinogens by their illicit abuse rather than by their responsible clinical use and apply these judgments to make decisions regarding its clinical use. One reason for this is that relatively few information is available to evaluate the safety and tolerability of psilocybin and other classical hallucinogens when administered to healthy human volunteers in controlled settings. Unfortunately, many of the older studies on adverse reactions are case reports or used small and unrepresentative samples. Furthermore, these studies were often poorly standardized and lacked adequate control groups and follow-up measures (Grob, 1994).

Thus, one further aim of the study presented in Chapter 3 was to improve knowledge about the risk-benefit ratio of psilocybin use in experimental psychopharmacology and to lay the foundations for future studies. To achieve this goal, the study reported about the proportions of subjects experiencing acute adverse reactions (so called "bad trips") under different dose conditions, various somatic and psychological complaints 24 h after drug administration, as well as longer lasting negative effects, such as Hallucinogen Persisting Perception Disorder (HPPD), prolonged psychosis, and negative changes in drug consumption habits. The latter effects were evaluated by analyzing a follow-up questionnaire that was completed 8-16 months after the last psilocybin session by 90 of 110 included subjects (82%).

1.4 ASSESSMENT OF ACUTE SUBJECTIVE EFFECTS

Despite considerable methodological advances in the area of neuroimaging, the neuronal correlates of consciousness can not be identified without data from the first person perspective. Consequently, investigators who use hallucinogens as tools to study the neural basis of ASCs must rely on well-validated rating scales to assess the subjective effects of these drugs. The revival of human hallucinogen research during the past 20 years has therefore greatly increased the demand for such ratings scales. Perhaps the currently most widely used instruments for assessing subjective effects of hallucinogens are the Hallucinogen Rating Scale (HRS) developed by Strassman, Qualls, Uhlenhuth, and Kellner (1994) and later validated by Riba, Rodríguez-Fornells, Strassman, and Barbanoj (2001) and the three versions of ASC rating scales (i.e., APZ, OAV, and 5D-ASC) developed by Dittrich and his colleagues (Dittrich, 1975b, 1985; Bodmer, 1989; Braun, 1997; Dittrich, 1998; Dittrich, Lamparter, & Maurer, 2010). The HRS and the three different versions of Dittrich's questionnaire have been administered in approximately 30 and 70 published experimental studies, respectively. Although both questionnaires have been translated into different languages, the HRS has been used predominantly

in the US and in Spain, while Dittrich's scales were most popular in Switzerland and in Germany. Vollenweider's research group has used Dittrich's rating scales in all its studies, and this probably has largely contributed to their popularity.

Compared to the HRS, Dittrich's questionnaires have the advantages that their primary factors were derived empirically by performing multivariate statistical analyses and that they were designed to be applicable to a wide range of ASCs, not just those induced by hallucinogenic drugs. The flexibility of Dittrich's questionnaires is not only reflected in their item content, but also in the fact that some of their scales are supposed to be measurement invariant across ASCs induction methods. Thus, they are probably more helpful to integrate research on ASCs than the HRS.

Because Dittrich's rating scales continue to be widely used and because many details of the most important psychometric investigations on these scales have not been published in English before, the historical development and validation of the three questionnaire versions APZ, OAV, and 5D-ASC is first summarized and then critically reviewed in Chapter 2 of this introduction. The review demonstrates that – from a modern methodological perspective – the existing psychometric studies have many methodological shortcomings and that some of these seriously weaken the conclusions drawn by Dittrich and his colleagues.

The review presented in Chapter 2 paves the way for the empirical study in Chapter 4, which analyzed the psychometric properties of the OAV in a relatively large sample of experimentally induced ASCs by implementing most of the suggestions for improvement made in Chapter 2. For instance, the study for the first time applied methods of the structural equation modeling (SEM) framework to one of Dittrich's questionnaire. These methods are associated with many of the methodological and statistical advances in quantitative psychology in the last two decades (Marsh, Lüdtke, et al., 2010) and are more appropriate for testing specific hypotheses on latent variable structures than the previously used exploratory methods.

1.5 PREDICTION OF PSILOCYBIN RESPONSE

Acute subjective experiences induced by hallucinogenic drugs differ considerably from person to person and even in the same individual on different occasions (Nichols, 2004). It is widely accepted that these differences largely result from differences in *set* (i.e. the preparation of the subject, his personality structure, and current mood state) and *setting* (i.e. the physical, social, and cultural environment in which the drug is taken) – terms that were originally coined by Timothy Leary (Leary, Litwin, & Metzner, 1963) and that soon after became a permanent part of the psychedelic idiom (Stafford & Golightly, 1967; Eisner, 1997).

The effects of hallucinogens are believed to be more strongly dependent on set and setting than those of other classes of drug (Nichols, 2004). Some authors have even gone so far as to claim that set and setting are by far the most important determinants of hallucinogenic drug effects. For instance, Leary (1964) wrote:

> Set and suggestive contexts account for ninety-nine percent of the specific response to the drug. Thus, you cannot sensibly talk about the effects of psilocybin. It's always the set and suggestive context triggered off by the drug. A fascinating tension between these two factors – set and context –

> inevitably develops. If both are positive and holy then a shatteringly sacred experience results. If both are negative then a hellish encounter ensues. There is, of course, the tendency for people to impose their familiar games on to the psilocybin experience. The more rigidly committed to the game, the stronger this tendency. If the drug-giving person is secure, flexible, supportive, then the experience is almost guaranteed to be pleasant and therapeutic. (p. 115)

Accordingly, Leary and others have considered set and setting as essential for the successful use of these drugs in psychotherapy and for personal growth. Moreover, they have put forward the ignorance of set and setting, particularly by people who use these drugs recreationally but also by some professionals, as one of the main explanations for the troubles that ultimately led to the banning of hallucinogens (Walsh & Grob, 2005). The importance of set and setting was recently reinforced by Griffiths, Richards, McCann, and Jesse (2006).

However, just like Leary and others have admonished the ignorance of set and setting, others have warned about their over-emphasis because it had led to a false sense of security. For instance Ungerleider, Fisher, Fuller, and Caldwell (1968) wrote:

> There is an ever-growing LSD mythology, too, much of it having to do with set and setting. For example, one commonly hears that a bad LSD experience will not result if:
> – One is in a calm frame of mind (no fights that day with spouse or employer);
> – One takes the LSD with one or two good friends or with an experienced sitter or guide present;
> – The room has soft lighting and a thick carpet or mattress to sit on;
> – One is listening to the Indian music of Ravi Shankar and reading reassuring phrases from the Tibetan Book of the Dead;
> – and perhaps if one has a "downer" or chlorpromazine pill at hand.
> But we have hospitalized many persons who had taken these precautions and who also had had up to 100 previous good LSD experiences. (p. 1489)

The above quotes clearly demonstrate that the evaluation of set and setting was strongly tied to questions about the safety and therapeutic usefulness of these drugs. Furthermore, it seems that those who were convinced of the benefits and safety of these drugs for political reasons tended to over-emphasize the importance of set and setting, while those who wanted the prohibit them tended to ignore or downplay these influencing factors. However, the incompatibility of the concepts of set and setting with the emerging orthodoxy that hallucinogenic drug effects were unpredictable and dangerous was probably not the only reason why influences of set and setting tended to be ignored. Langlitz (in press) points out:

> For scientific, disciplinary, economic, and political reasons, biological psychiatry and psychopharmacology had an interest in attributing the effects of drugs to the drugs alone. This ideology of "pharmacologicalism" helped psychiatry to be acknowledged as part of scientific medicine, enabled pharmaceutical companies to fulfill the Food and Drug Administration's regula-

> tory requirement to demonstrate specificity of drug action, and legitimized the War on Drugs (DeGrandpre, 2006). (p. 259)

To escape these ideological traps, it is decisive to evaluate the importance of set and setting on strictly empirical grounds. Unfortunately, only few empirical studies have investigated the determination of hallucinogen response by set and setting and most of these were conducted in the 1950s and 1960s when methodological standards were lower than today. For instance, many of the these studies used poorly validated measures and relied on small sample sizes. Furthermore, except for one newer study conducted by Dittrich and his colleagues (Dittrich, 1994; Dittrich & Lamparter, 1994), they only obtained a few predictor variables at a time and/or did not adjust for confounding variables. Hence, they could not provide information on the order of importance of different predictor variables from a wide range of different domains. The lack of high quality studies in this area is unfortunate because knowledge about set and setting could improve the safety and standardization of controlled experiments and significantly advance our understanding of the neurobiological systems involved in the actions of hallucinogens.

Thus, the study presented in Chapter 5 aimed to predict various dimensions of psilocybin response by analyzing 24 potentially important predictor variables and by using a much larger sample size (n = 409) and more sophisticated statistical methods than in previous studies. The analyzed predictors covered a wide range of domains and captured both inter- (i.e., stable personality traits, demographic variables, drug use and pre-experiences) and intraindividual differences (i.e., mood state immediately before drug intake and psychological problems during the last four weeks before drug intake), as well as external influences (i.e. whether the experimental session involved PET measurements or not, time of measuring drug response, and drug dose).

2

ON THE ASSESSMENT OF ALTERED STATES OF CONSCIOUSNESS BY DITTRICH'S APZ, OAV, AND 5D-ASC QUESTIONNAIRES: A CRITICAL REVIEW

2.1 INTRODUCTION

Dittrich's APZ questionnaire and its revised versions, OAV and 5D-ASC, are among the most widely used self-report questionnaires for assessing subjective experiences of altered states of consciousness (ASCs). The use of these questionnaires has been most popular in psychopharmacological research. However, despite their widespread application, few studies have investigated the psychometric properties of these rating scales.

This chapter is divided into two parts. In the first part, the available literature on the development and validation of the APZ, OAV, and 5D-ASC questionnaires is critically reviewed. Although the validation studies of the APZ have been summarized elsewhere (Dittrich, von Arx, & Staub, 1985; Dittrich, 1998), detailed information on the OAV and 5D-ASC questionnaires has only appeared in unpublished master and doctoral theses (Bodmer, 1989; Braun, 1997; Habermeyer, 1999) or in hard to come by book chapters (Bodmer, Dittrich, & Lamparter, 1994), all of which are written in German. Thus, a great deal of the information presented in this chapter has not been available to the English-speaking world before.

The second part of this chapter discusses methodological shortcomings of the existing APZ, OAV, and 5D-ASC validation studies from a modern methodological perspective. Due to substantial methodological advances in the past 20 years, particularly in the area of latent variable modeling, many of the methodological problems highlighted in this article have not been recognized previously. I will discuss the implications of these methodological shortcomings with respect to the interpretation of previously obtained results. Furthermore, I will propose alternative statistical procedures that should be used in future validation studies and that may lead to a revision of the scales.

2.2 QUESTIONNAIRE DEVELOPMENT AND VALIDATION

2.2.1 *The APZ Questionnaire*

2.2.1.1 *Scale construction*

Many authors have described strong similarities between different deliberately induced ASCs (e.g., by hallucinogenic drugs, hypnosis, meditation, sensory deprivation) and disease-induced ASCs (e.g., schizophrenic psychosis) leading them to hypothesize that ASCs, independent of their means of induction, have a common core (Dittrich, 1985, 1996). In the western literature, the hypothesis of a common core of ASCs was put forward more than 165 years ago by Moreau de Tours (1845), but more or less explicit formulations of this hypothesis can also be found in the works of Masters and Houston

(1967), Huxley (1959), Weil (1972), and especially of Ludwig (1966). Dittrich (1985, 1996) was the first to formulate this hypothesis in such a way that it can be tested empirically:

> Irrespective of their mode of induction, ASCs have invariant features in common, which at the same time differentiate them from normal waking consciousness. These etiology-independent characteristics form a structure of mutual similarities, which is maintained when ASCs are induced by different means. On the dimensional level (i.e., methodologically adopting an approach similar to that of dimensional theories of personality) this means that ASCs have certain major dimensions in common, irrespective of their induction means and intensity. Of course, this does not exclude that etiology-specific dimensions exist, such as probably, "clouding of consciousness" or "vigilance reduction" for hallucinogens of the second order. (Dittrich, 1998, p. 81)

Dittrich (1985, 1996) reasoned that, if the above hypotheses could not be falsified for a broad range of ASC induction methods, integration of phenomenological, psychophysiological, and neurobiological research on ASCs would be greatly enhanced. Not only would it lay the foundation for a more coherent definition of the term ASC, but the common features of ASCs could also be reduced to common latent dimensions and thus described more parsimoniously. Furthermore, because these common latent dimensions would be invariant across ASC induction methods, different ASCs could be directly compared by their scores on these dimensions. This, in turn, would eventually lead to an empirical taxonomy of ASCs.

Dittrich (1985, 1996) tested the above hypotheses in a series of experimental studies, in which ASCs were induced in healthy subjects by eleven different induction methods (n = 259) or control condition procedures (n = 134). The studied induction methods were divided into four groups:

1. Hallucinogens of the first order (*N*,*N*-dimethyltryptamine [DMT], psilocybin, and Δ^9-tetrahydrocannabinol [THC]; n = 82)
2. Hallucinogens of the second order (nitrous oxide; n = 38)
3. Sensory deprivation in a broader sense (perceptual deprivation, hypnagogic states, autogenic training, and hypnosis; n = 79)
4. Sensory overload (stimuli of high variety; n = 60)

The APZ questionnaire, which was originally constructed by Dittrich (1975b) to assess deliberately-induced as well as disease-induced ASCs, served as the primary outcome measure in these experiments. The items of the APZ were derived from previously existing questionnaires on ASCs, narrative reports, psychiatric rating scales, and the author's personal experience with ASCs. They were chosen to cover a broad range of phenomena potentially occurring during ASCs. Each item consists of a statement describing a specific experience of ASC in the first person singular and past tense. By responding with "yes" or "no", subjects are forced to either confirm or negate these experiences.

From the 158 items of the APZ questionnaire, Dittrich (1985, 1996) identified 72 items meeting his criteria of etiology-independency. That is, the lower bounds of the 95% confidence intervals of proportions of subjects answering these items with "yes" were bigger than 0.01 in each of the four main groups of stimuli, and the items were also significantly differentiating ASC from normal waking consciousness (p <

0.05, comparisons with control groups and the pre-experimental level in the combined groups).

By analyzing the correlation matrices of the 72 etiology-independent items using exploratory factor and cluster analysis and based on considerations of stability, reliability, and interpretability, Dittrich determined three primary and one secondary etiology-independent dimensions. The three primary dimensions were termed Oceanic Boundlessness (OBN), Dread of Ego Dissolution (DED), and Visionary Restructuralization (VRS) and, according to Dittrich (1985, 1996), operationally define what the writer Aldous Huxley (1959) has described as heaven, hell, and visions. The name of the OBN scale was derived from a psychological term that was originally coined 1927 in a letter from Romain Rolland to Sigmund Freud (Rolland, 1967) and later popularized by Sigmund Freud in his book *Civilization and Its Discontents* (1930/1961). The OBN scale includes items measuring positively experienced depersonalization and derealization, ecstatic experiences of bliss, and feelings of eternity and unity. High scores on the OBN scale therefore indicate a state similar to mystical experiences as described in the scientific literature on the psychology of religion (e.g. Stace, 1961). The DED scale includes items measuring negatively experienced derealization and depersonalization, cognitive disturbances, catatonic symptoms, paranoia, and loss of thought and body control. High scores on the DED scale therefore indicate a very unpleasant state, similar to so called bad or horror trips described by drug-users. The VRS scale contains items measuring visual (pseudo)-hallucinations, illusions, auditory-visual synesthesiae, and changes in the meaning of percepts. The secondary scale Global Altered States of Consciouness (G-ASC) consists of the 72 etiology-independent items and can be interpreted as a general measure of consciousness alteration.

2.2.1.2 *Re-examinations of the factorial structure*

In order to test the external validity of the experimental results in the field and to construct psychometrically equivalent scales in several languages, the International Study on Altered States of Consciousness (ISASC) was carried out (Dittrich, von Arx, & Staub, 1985). In this study, 1133 subjects from six different countries (i.e., Switzerland, Germany, Italy, Great Britain, Portugal, and USA) and four different languages (i.e., German, Italian, English, and Portuguese) described the most recent ASC they had experienced within the past 12 months using different language versions of the APZ. The described ASCs were most frequently induced by cannabis (44%), LSD (8.5%), meditation (7.6%), hypnagogia (5.6%), listening to music (4.5%), solitude and isolation (4.0%), autogenic training (2.6%), and hypnosis (1.3%). In 70% of the cases, ASCs were induced by a drug, which is a significantly higher percentage than in the experimental studies (46.3%). An analysis of the similarity of the correlations matrices and factor solutions formed from seven different data sets (experiments and field data from six different countries) yielded results that corresponded well to those of the experimental studies. The authors concluded that the external validity of the experiments was demonstrated in spite of important differences between the two studies and that the four studied language versions of the APZ are psychometrically equivalent (Dittrich, von Arx, & Staub, 1985; Dittrich, 1998). The APZ was also translated into Dutch, Finnish, French, Greek, Spanish, and Russian, but the psychometrical equivalence of these versions has not yet been tested (Dittrich, 1998).

Two additional studies analyzing the factorial structure of the APZ have been published. In the first study (Gouzoulis-Mayfrank, Habermeyer, et al., 1998), 93 hospitalized schizophrenic patients were asked to describe the acute phase of their last psychotic episode using the APZ. The 72 etiology-independent items of the completed APZ questionnaires were analyzed by a principal component analysis (PCA) with three factors and varimax rotation. The authors reported that only 56.3% of items were correctly distributed to their hypothesized factor. The proportion of correctly distributed items was higher for the DED factor (61.9%) than for the OBN (53.8%) and VRS factors (50.0%). Although the hypothesized factorial structure was poorly reproduced in this study, it should be noted that the definition of ASC underlying the APZ scale construction did not include ASCs induced by psychiatric diseases (e.g. see Dittrich, 1998). Hence, it might be argued that the studied sample was not suitable to falsify the factorial structure of the APZ. Further shortcomings of the study are the low sample size and the use of PCA instead of principal factor analysis. In the second study (Bodmer, 1999), the re-examination of the factorial structure of the APZ was based on a sample of 135 subjects in which ASCs were experimentally induced by either DMT, nitrous oxide, or sensory deprivation. After performing a principal factor analysis with three factors and orthogonally rotating the loading matrix to maximal similarity with the hypothesized factorial structure, 74% of items were correctly distributed to their hypothesized factor. The disagreement of the empirically derived and hypothesized factorial solutions in this study mainly resulted from a number of undifferentiated items having salient loadings on the OBN factor and DED items having no salient loadings on any factor.

2.2.1.3 *Reliability assessments*

The reliabilities of the APZ scales have been estimated by the Kruder-Richardson formulas KR-8 and KR-20. The KR-20 formula gives identical reliability estimates as Cronbach's α when applied to dichotomous variables (Thompson, 2003). In the experimental studies (Dittrich, 1985, 1996), the OBN, DED, and VRS scales showed satisfactory reliabilities, while the reliability of the G-ASC scale was excellent. In the field studies (Dittrich, von Arx, & Staub, 1985), the reliability of the OBN scale was substantially reduced, while all other scales showed similar reliabilities. The reliability of the OBN scale was unsatisfactory ($r < 0.8$) in four of six countries. Test-retest reliabilities calculated on the basis of repeated measurements approximately one year apart were reported for a small sample of pharmacologically induced ASCs ($n = 61$ Dittrich, 1985, 1996). The estimated retest-reliabilities in this sample ranged from 0.56 for the DED scale to 0.71 for the OBN scale.

2.2.1.4 *Validity assessments*

Convergent validities of the APZ scales have been assessed by correlating them with the scales of the DAE, a questionnaire measuring subjective effects of cannabis (Dittrich, Baettig, Woggon, & Zeppelin, 1972), and with five scores of cognitive performance tests. As predicted, most correlations were statistically significant and of at least medium effect sizes (Dittrich, 1985, 1996).

Discriminant validities of the APZ scales have been assessed in two ways. First, because the APZ was designed to assess specific states, but not traits, the APZ scales

were correlated with the scales of the AUPI personality questionnaire (U. Baumann & Dittrich, 1975; U. Baumann & Dittrich, 1976). As predicted, most correlations were low. The highest correlations were found between DED and Psychoticism (r = 0.19) and DED and Neuroticism (r = 0.19; Dittrich, 1985, 1996). Second, because the APZ was supposed to measure dimensions of ASCs, but not mood changes in general, a double-blind study was performed in which the effects of ethyl alcohol, chlorpromazine, and placebo on the scale scores of the APZ were compared. As predicted, the APZ scales did not differentiate between the three experimental conditions (Dittrich, 1985, 1996). To the author's knowledge, only one further study has reported correlations between the APZ scales and other psychological constructs. Specifically, Gouzoulis-Mayfrank, Habermeyer, et al. (1998) computed correlations between the APZ and the Brief Psychiatric Rating Scale (BPRS) in a sample of 93 endogenously psychotic patients. They found statistical significant correlations of medium to large effect sizes between OBN and BPRS Thought Disorder (r = 0.33) and between DED and BPRS Anxiety/Depression (r = 0.42).

2.2.1.5 *Applications*

According to Dittrich (1998, pp. 84), "the APZ questionnaire has become the standard instrument when assessing the common denominator of ASCs." A literature search reveals that approximately 25 studies using the APZ have been published. Although the APZ was developed to measure a broad spectrum of ASCs, the APZ has been used predominantly in psychopharmacological research. Pharmacological inducers of ASCs that have been investigated by the APZ include psilocybin (Dittrich, 1985; Vollenweider, Leenders, Scharfetter, P. Maguire, et al., 1997; Gouzoulis-Mayfrank, Thelen, et al., 1999; Griffiths, Richards, McCann, & Jesse, 2006; Griffiths, Johnson, et al., 2011; Maclean et al., 2011), ketamine (Bolle, 1988; Vollenweider, Leenders, Scharfetter, Antonini, et al., 1997), DMT (Bickel, Dittrich, & Schöpf, 1976; Dittrich, 1985, 1994), ayahuasca (Riba, Rodríguez-Fornells, & Barbanoj, 2002), Salvia divinorum (González, Riba, Bouso, Gómez-Jarabo, & Barbanoj, 2006), Salvinorin A (Mendelson et al., 2011), mescaline (Hermle, Fünfgeld, et al., 1992), THC (Dittrich, Bickel, & Zimmer, 1975; Dittrich, 1975a; Dittrich, Bickel, Schöpf, & Zimmer, 1976; Koukkou & Lehmann, 1976, 1978; Dittrich, 1985), nitrous oxide (Dittrich, 1985, 1994), ethanol (Dittrich, 1985), chlorpromazine (Dittrich, 1985), methylphenidate (Griffiths, Richards, McCann, & Jesse, 2006), 3,4-methylenedioxy-N-ethylamphetamine (MDE; Hermle, Spitzer, Borchardt, Kovar, & Gouzoulis, 1993; Gouzoulis-Mayfrank, Thelen, et al., 1999), d-amphetamine (Vollenweider, R. P. Maguire, Leenders, Mathys, & Angst, 1998), and d-methamphetamine (Gouzoulis-Mayfrank, Schreckenberger, et al., 1999). Non-pharmacological means of induction investigated by APZ include sensory deprivation (Dittrich, 1975a, 1985, 1994), hypnagogic states (Dittrich, 1985), autogenic training (Würmle, 1977; Dittrich, 1985), hypnosis (Dittrich, 1985), sensory overload (Dittrich, 1985), mind machines (Walach & Käseberg, 1998), endogenous psychosis (Gouzoulis-Mayfrank, Habermeyer, et al., 1998), and sweat lodges (Polito, Langdon, & Brown, 2010). A shortened version of the APZ was also used to assess ASCs induced by childbirth (Gruzdev & Spivak, 2006).

2.2.2 *The OAV Questionnaire*

2.2.2.1 *Scale construction*

Although reliabilities and validities of the APZ scales were deemed to be acceptable in the experimental as well as in the field studies, several weaknesses were also recognized. For example, because the transition from normal waking consciousness to ASCs is continuous rather than abrupt, the binary item response format of the APZ was too crude to measure subtle alterations of consciousness. Consequently, several APZ items showed very low endorsement rates when measuring experiences of ASCs of low intensity, which in turn attenuated the item discriminations and scale reliabilities. Other recognized shortcomings of the APZ concerned the relatively low number of items assigned to the OBN and VRS dimensions, the low conceptual breadth of the VRS dimension, as well as problematic wordings of several items. A psychometrically improved version, named OAV, was therefore developed by Bodmer (1989) in the context a of diploma thesis carried out at the Department of Psychology at the University of Zürich. Its name is derived from the first letters of the German names of the primary three dimensions of the APZ (i.e., **A**ngstvolle Ichauflösung, **O**zeanische Selbstentgrenzung, **V**isionäre Umstrukturierung). Because the OAV was supposed to measure the primary three dimensions of the APZ only, its item pool is primarily based on the 72 etiology-independent items of the APZ. However, the wordings of several items were changed, some new items were introduced, and some items were completely dropped. The reformulation of items aimed not only at reducing cross-loadings, decreasing ambiguity, and enhancing ease of understanding, but also at widening the conceptual breadth of the OBN and VRS dimensions.

Because the analyses of the APZ had indicated that the VRS dimension describes not only changes in visual perceptions and their associated meanings, but also a general increase in the perception of internally produced stimuli, Bodmer (1989) hypothesized that the VRS dimension could be conceptually extended by incorporating items that measure an increase of imaginations, associations, and memory retrieval. The proposed re-conceptualization of the VRS dimension was mainly based on theoretical considerations of Leuner (1962, 1981), who had speculated that hallucinogenic drugs elicit visual hallucinations by intensifying internal imagery such that the distinction between internally produced images and external perceptions becomes blurred. Furthermore, because previous analyses of the APZ had indicated that high scores on the OBN dimension describe a state similar to mystical experiences, new items for the OBN scale were formulated on the basis of six of the nine categories of mystical experiences proposed by Stace (1961). Hence, it was aimed to shift the conceptualization of the OBN scale toward a more focused and complete assessment of mystical experiences.

In order to allow a more fine-grained assessment of subjective experiences of ASCs and to reduce item difficulties, the item response format was changed from binary to visual analogue. Unlike the APZ items, which allowed only "yes" and "no" answers and thus, according to the response format typology of Russell and J. M. Carroll (1999), were "ambiguous unipolar", the visual analogue scales of the OAV were defined in a "strictly unipolar" manner. That is, the intensity of subjective experiences of ASCs were measured along a continuum that goes from *no, not more than usual* on the left to *yes, very much more than usual* on the right. By repeating these new descriptions of item

end poles for every item, the author aimed to ensure that only non-ordinary states of consciousness were assessed by the questionnaire.

The first draft of the OAV questionnaire contained 70 items and was validated in field study, in which 177 subjects retrospectively rated an ASC they had experienced in the past 12 months. Similar to the field studies using the APZ, the majority of the described ASCs (74.6%) were induced by psychotropic substances. The factorial structure of the OAV in the collected data set was examined by conducting a principal factor analysis and orthogonally rotating the factor matrix to maximal similarity with the hypothesized factor matrix. Taking the highest loadings as decision criterion, it was determined that only four items were not assigned to their hypothesized factor. After excluding these four items, the final version of the OAV contained 66 items in total, of which 27, 21, and 18 were assigned to the OBN, DED, and VRS dimension, respectively.

The validation study of Bodmer (1989) indicated that the questionnaire revision successfully improved several psychometric properties. Specifically, compared to the original APZ questionnaire, the OAV showed lower item difficulties, higher item discriminations, improved simple structure, and higher scale reliabilities. The improved scale reliabilities of the OAV could not be explained by the increased scale lengths alone because a comparison between the APZ and OAV scale reliabilities corrected for different scale length by the Spearman-Brown formula still showed superior scale reliabilities in the OAV. Because 49 subjects of the validation study had completed both the APZ and the OAV questionnaire, similarities of the factors were assessed by simply correlating the raw summated scales of each of the two questionnaires. As expected, the DED scale, which had been revised the least, showed the highest correlation (r_{DED} = 0.87) between the two questionnaires. However, the correlation coefficient for the VRS scales (r_{VRS} = 0.85) was almost equally high and substantially higher than the correlation of OBN scales (r_{OBN} = 0.58). Given that the conception of the VRS dimension had been revised more than the OBN dimension, this finding was rather surprising.

A later study (Bodmer et al., 1994), which used a larger sample (n = 184), corroborated the finding that the VRS scale correlates higher than the OBN scale across the two questionnaires. The lower correlation of the OBN scale might be explained by the facts that the OAV-OBN scale contained the lowest percentage of unchanged items (19%) and that the length of the scale was increased the most, namely, from 13 items in the APZ to 27 items in the OAV. However, simple correlations of summated scales must be interpreted very cautiously because they are biased estimates of the true factor correlations in at least three ways. First, they are attenuated by the measurement errors of the two scales. Second, they do not take (non-target) cross-loadings into account. Unless all the cross-loadings are close to zero, this inflates the correlations of summated scales (e.g. see Marsh, B. O. Muthén, et al., 2009). Third, because the two questionnaires were completed by the same subjects, responses to the same items are typically more positively correlated than can be explained in terms of correlations between the factors they represent. Unless this type of measurement error is controlled by a confirmatory factor analysis (CFA) model that includes correlated uniquenesses, the estimated factor correlations are positively biased (Marsh & Hau, 1996).

2.2.2.2 *Re-examinations of the factorial structure*

The hypothesized re-conceptualization of the VRS dimension was not rejected in the original OAV validation study of Bodmer (1989). That is, the newly constructed items assessing increased imagination and memory retrieval indeed loaded highest on the VRS factor. However, in a later study (Bodmer, 1999), in which the factorial structure of the OAV was re-examined by analyzing a sample of 135 experimentally induced ASCs (for a description of the sample, see Dittrich, 1994), the VRS factor was poorly recovered. Specifically, only 5 of the 18 VRS items loaded highest on the VRS factor, while all remaining VRS items loaded highest on the OBN factor. The correctly assigned VRS items exclusively measured aspects of elementary visual hallucinations and auditory-visual synesthesiae. Hence, the distribution of items in this sample did not support the hypothesis that a single latent construct accounts for the correlations between alterations in visual perception, changed meaning of percepts, increased imagination, and memory retrieval.

The hypothesized structure of the VRS dimension introduced by Bodmer (1989) also poorly fitted the data in a sample of 93 schizophrenic patients who described their last acute psychotic episode by the OAV questionnaire (Gouzoulis-Mayfrank, Habermeyer, et al., 1998; Habermeyer, 1999). Although a principal component analysis with three factors and subsequent varimax rotation correctly assigned 88.9% of OBN and 95.2% of DED items to their hypothesized factors, this was only true for 44.4% of the VRS items. As in the sample of experimentally induced ASCs (Bodmer, 1999), the VRS dimension was reduced to a core of items measuring aspects of elementary visual hallucinations and auditory-visual synesthesiae, whereas VRS items measuring changed meaning of visual percepts, increased imagination, and memory retrieval loaded highest on the OBN factor. Despite these similar patterns of misassignments, the authors of both studies concluded that the configurations of item loadings were in sufficient agreement with the original hypotheses.

2.2.2.3 *Reliability assessments*

The reliabilities of the OBN scales have been estimated by the variance analytic approach of Hoyt (1941), which produces identical results as Cronbach's α (Thompson, 2003). All four scales of the OAV demonstrated high to excellent reliabilities in the field (n = 177; Bodmer, 1989) and in the experiments (n = 135; Bodmer, 1999). With the exception of the VRS scale in the field data set having a reliability of 0.89, all reported OAV scale reliabilities were above 0.9.

Test-retest reliabilities were calculated on the basis of 112 subjects completing the OAV twice (i.e., immediately after the experiments and after one year). The values ranged from 0.77 for the VRS scale to 0.83 for the OBN scale (Bodmer, 1989).

2.2.2.4 *Validity assessments*

The validity of the OAV has been deduced from its high correlation with the APZ and from its proved value in discriminating the effects of different ASC inducing agents. To the author's knowledge, correlations between OAV scales and other psychological constructs have been reported in only one study. Specifically, in a sample of 93 schizophrenic patients who described their last acute psychotic episode by the OAV

questionnaire (Gouzoulis-Mayfrank, Habermeyer, et al., 1998), OBN and G-ASC were significantly correlated with the BPRS Thought Disturbance scale (r = 0.41 and r = 0.33, respectively) and DED was significantly correlated with the BPRS Anxiety-Depression scale (r = 0.5).

2.2.2.5 *Applications*

Similar to the APZ, the OAV has been used predominantly in psychopharmacological research. At least 25 studies using the OAV have been published. A large proportion of these studies involved the administration of 3,4-methylenedioxy-methamphetamine (MDMA; Vollenweider, Gamma, Liechti, & Huber, 1998; Vollenweider, Remensberger, Hell, & Geyer, 1999; Gamma et al., 2000; Liechti & Vollenweider, 2000; Liechti, C. Baumann, Gamma, & Vollenweider, 2000; Liechti, Saur, Gamma, Hell, & Vollenweider, 2000; Frei et al., 2001; Liechti, Geyer, Hell, & Vollenweider, 2001; Liechti, Gamma, & Vollenweider, 2001; Liechti & Vollenweider, 2001; Vollenweider, Liechti, & Paulus, 2005), psilocybin (Vollenweider, Vollenweider-Scherpenhuyzen, et al., 1998; Vollenweider, Vontobel, Hell, & Leenders, 1999; Hasler, Bourquin, Brenneisen, & Vollenweider, 2002; Quednow, Kometer, et al., 2012), and ketamine (Vollenweider, Vontobel, Oye, Hell, & Leenders, 2000; Umbricht, Koller, et al., 2002; Gouzoulis-Mayfrank, Heekeren, Neukirch, et al., 2005; Daumann, Heekeren, et al., 2008; Daumann, Wagner, et al., 2010), but some studies also used the OAV to measure the subjective effects of DMT (Gouzoulis-Mayfrank, Heekeren, Neukirch, et al., 2005; Daumann, Heekeren, et al., 2008; Daumann, Wagner, et al., 2010), MDE (Spitzer, B. Franke, et al., 2001), and 3,4-methylenedioxy-amphetamine (MDA; Baggott et al., 2010). To the author's knowledge, only one study has used the OAV to measure non-pharmacologically induced ASCs, namely, ASCs induced by schizophrenia (Gouzoulis-Mayfrank, Habermeyer, et al., 1998).

2.2.3 *The 5D-ASC Questionnaire*

2.2.3.1 *Scale construction*

Although the dimensional analyses of the APZ and OAV questionnaires had revealed three "etiology-independent" dimensions of ASCs, Dittrich's own investigations as well as the scientific literature on ASCs (e.g. Vaitl et al., 2005) pointed to the existence of further dimensions that are specific to certain ASC-inducing agents. For example, acoustic alterations and hallucinations are a common feature of non-ordinary states of consciousness induced by certain psychiatric diseases, such as schizophrenia and alcohol withdrawal psychoses, and have been described under conditions of sensory deprivation and hypnagogic states (Mavromatis, 1987), but seem to be less common in hallucinogen-induced ASCs (Malitz, Wilkens, & Esecover, 1962). In accordance with these findings, only 2 of the 11 APZ items measuring acoustic alterations met criteria of etiology-independency in the experimental studies described above (Dittrich, 1985; Schneiter, 1991). Furthermore, clouding of consciousness and reduction of vigilance are characteristic features of hallucinogens of the second order and of sedative drugs, but not of hallucinogens of the first order (Leuner, 1981). Dittrich therefore hypothesized that "auditory alterations" and "reduction of vigilance" were two etiology-dependent

dimensions, which, in addition to the three primary etiology-independent dimensions, could be reliably and validly measured.

In order to test these hypotheses, Dittrich and co-workers constructed the Bewusstseinstrübung und Akustische Halluzinationen (BETA) questionnaire (Dittrich, Lamparter, & Maurer, 1989; Schneiter, 1991), which contains 17 items measuring Vigilance Reduction (VIR) and 22 items measuring Auditory Alterations (AUA). The AUA dimension was conceptualized analogously to the VRS dimension. That is, the dimension was not only supposed to measure auditory (pseudo-)hallucinations, but also changes in the meaning of auditory perceptions and a general increase of internally produced auditory stimuli. To reduce item difficulty, items measuring auditory hallucinations were formulated such that they basically characterized non-psychotic hallucinations. That is, in addition to a description of a specific auditory hallucination, every item also stated that the deceptive character of the auditory percept was noticed or at least suspected. Analogous to the VRS dimension, the AUA dimension was conceptualized to cover hallucinations of a broad range of vividness and complexity. Specifically, five items were constructed to measure amorphic hallucinations (e.g., buzzes, drones and soughs), three items were constructed to measure low-structured hallucinations (e.g., rattles, ticks, rings) and six items were constructed to measure high-structured hallucinations (e.g., sounds, melodies, words, sentences). The VIR dimension was designed to cover experiences of sleepiness, clouding of consciousness and perceptions, and slowing down of thoughts and actions. Similarly to the OAV questionnaire, the BETA contained a strictly unipolar visual analogue item response format.

In the context of a diploma thesis carried out at the Department of Psychology at the University of Zürich, Braun (1997) investigated whether the two hypothesized dimensions of the BETA questionnaire could be reliably and validly measured and whether they could be differentiated from the primary three dimensions of the APZ questionnaire. Braun (1997) pooled APZ and BETA questionnaire data from a study in which ASCs were experimentally induced (n = 135) and from two field studies (n = 118). She then performed a dimensional analysis on the Pearson correlation matrix of the 39 BETA items and the 49 APZ items comprising OBN, DED, and VRS scales. Because the BETA items were measured on a visual analogue scale and the APZ items on a binary scale, BETA items were dichotomized at their median prior to the computation of the Pearson correlation matrix. After conducting a principal axis factor analysis with five factors and a varimax rotation, a factorial solution emerged that was deemed to be in good agreement with the hypothesized structure. Taking the highest loadings as decision criterion, 80.7% of items were correctly distributed to their hypothesized factor. However, an item cluster analysis using the k-means algorithm and the Euclidian distance metric only recovered the DED, VIR, and VRS factors, whereas the OBN and AUA factors were merged into the fourth cluster, and a very small cluster containing two VRS and one OBN items emerged as a fifth cluster. Although, only 69% of all items were correctly assigned to their hypothesized factor in the cluster analysis, an assessment of the item distribution similarities by Cohen's $\varkappa$ indicated that the hypothetical and empirically derived solutions corresponded to a sufficient degree. Braun (1997) therefore concluded that the hypothesized five-dimensional structure was not falsified.

Based on the results of the exploratory factor analysis (EFA), eleven items were excluded from the AUA and VIR scales because they either loaded highest on the wrong

factor or violated criteria of simple structure. The combination of the original OBN (13 items), DED (22 items), and VRS (14 items) scales from the APZ with the revised AUA (16 items) and VIR (12 items) scales from the BETA resulted in a new questionnaire containing 77 items that measure five dimensions of ASCs. The new questionnaire demonstrated satisfactory scale reliabilities and item discriminations in the sample of the validation study. By comparing the means of the scale scores of six ASC-induction techniques (sensory deprivation, DMT, N_2O, LSD, Cannabis, and Opiates) using univariate analyses of variance, Braun (1997) also demonstrated that the five scales differentiated well between stimulus conditions.

The validation study of Braun (1997) eventually led to the publication of the Five Dimensions of Altered States of Consciousness (5D-ASC) questionnaire (Dittrich, Lamparter, & Maurer, 1999), which combines all items of the OAV with the revised AUA and VIR scales of the BETA. The 5D-ASC thus contains 94 items of which 27, 21, 18, 16, and 12 items are assigned to the OBN, DED, VRS, AUA, and VIR dimensions, respectively. The 5D-ASC uses the same item response format as the OAV questionnaire. Because the description of item end poles in the BETA is slightly different from that of the OAV, this means that the item response format of the AUA and VIR items slightly changed, when they were incorporated into the 5D-ASC. The historical development of Dittrich's ASC questionnaires is summarized in Figure 1.

2.2.3.2 *Reliability and validity assessment*

Because the OBN, DED, VRS, G-ASC scales of the 5D-ASC contain the same items as those of the OAV, the reliabilities and validities of these scales have been derived from the OAV validation studies described above. The reliabilities of the AUA and VIR scales of the 5D-ASC have been derived from studies using the BETA questionnaire. In a pooled data set derived from these studies (n = 253), both scales demonstrated reasonably high reliabilities when estimated by Cronbach's α, that is, 0.93 and 0.86 for the AUA and VIR scales, respectively (Braun, 1997). To the author's knowledge, correlations between the 5D-ASC scales and other psychological constructs have not yet been reported.

2.2.3.3 *Applications*

The 5D-ASC questionnaires has been used in at least 18 published studies. Similar to its predecessors, the 5D-ASC has been most popular in psychopharmacological research. Among the drugs most often studied by the 5D-ASC are psilocybin (Hasler, Grimberg, et al., 2004; Carter, Burr, et al., 2005; Carter, Pettigrew, Hasler, et al., 2005; Vollenweider, Csomor, et al., 2007; Wittmann et al., 2007; Wackermann et al., 2008; Kometer et al., 2011; Carhart-Harris, Williams, et al., 2011; Grob et al., 2011; Schmidt, Bachmann, et al., 2011), ketamine (Passie, Karst, et al., 2003; Northoff et al., 2005; Sprenger et al., 2006; Musso et al., 2011; Schmidt, Bachmann, et al., 2011), and MDMA (Hasler, Studerus, Lindner, Ludewig, & Vollenweider, 2009; Hysek, Simmler, et al., 2011; Hysek, Brugger, et al., 2012). Non-pharmacologically induced ASCs investigated by the 5D-ASC include sound-induced trance states (Fachner & Rittner, 2004; Hübner, 2007). A shortened and modified version of the 5D-ASC, called Experience of Deviation from Normal state (EDN), was used to assess states of relaxation induced by sensory deprivation in flotation tanks (Kjellgren, Sundequist, Norlander, & Archer, 2001; Kjellgren, Sundequist, Sundholm,

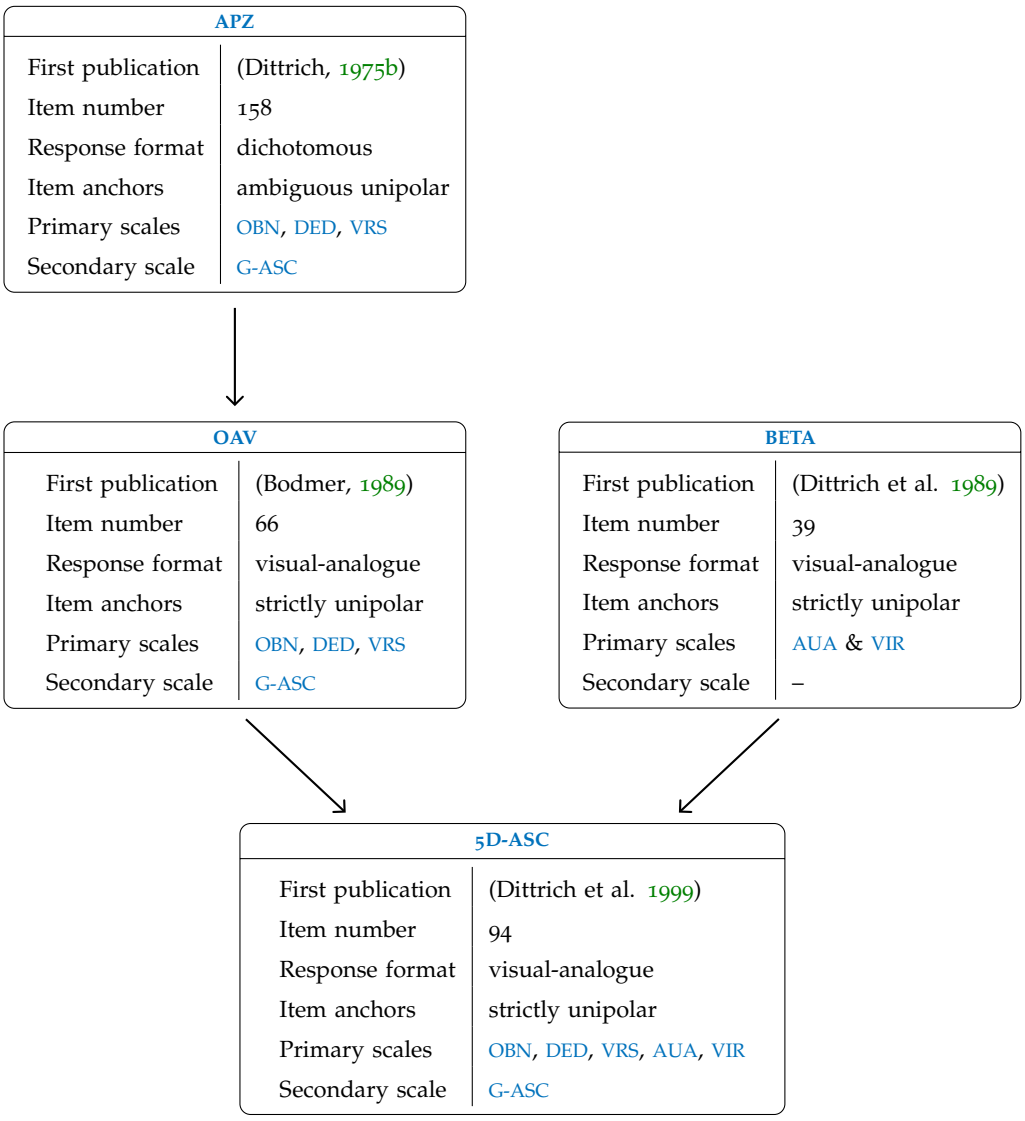

Figure 1: Dittrich's Altered States of Consciousness Rating Scales

Norlander, & Archer, 2004), Yogic breathing (Kjellgren, Bood, Axelsson, Norlander, & Saatcioglu, 2007), natural and simulated natural environments (Kjellgren & Buhrkall, 2010), and spike mats (Kjellgren, Erdefelt, Werngren, & Norlander, 2011).

2.3 METHODOLOGICAL ISSUES

Although Dittrich (1998) concluded that his original hypotheses on ASCs have survived considerable falsification testing not only in experimental but also in field studies and that the APZ questionnaire has become a psychometrically well-validated instrument for the assessment of "etiology-independent" features of ASCs in a "etiology-independent" three-dimensional space, I will demonstrate in this section that the methodological problems of the existing studies are severe enough to cast doubt on such conclusions. I will particularly address problems associated with the item selection, dimensional

analyses, measurement invariance evaluations, binary items, and reliability analyses, as these – in my opinion – are the biggest cause of concern.

2.3.1 *Item Selection*

According to Goldberg and Velicer (2006, pp. 212), the item selection "is by far the single most important decision to be made in any investigation, and it should be guided by theory and/or the findings from past research". This is because the item selection will strongly determine the resulting factor structure (e.g. see Velicer & Fava, 1998). Unfortunately, because little empirical research existed on the dimensionality of ASCs at the time of the APZ questionnaire development, the selection of APZ items was not guided by strong theory. Hence, it could not be assured that all important dimensions and/or facets of ASCs were covered by an equal number of items and by an equal conceptual breadth. Although the 158 items of the APZ have been formulated on the basis of previously existing questionnaires on ASCs, narrative reports, psychiatric rating scales, and the author's personal experience with ASCs, it remains questionable, whether the selected APZ items and, consequently, the resulting factorial structure were representative for the domain of interest. The confidence in the proposed factorial structure could have been improved by repeating the original dimensional analyses with other sets of items that are independently sampled from the domain of interest. However, this has not yet been attempted. The confirmation of the three-dimensional structure with the items sets of the OAV and 5D-ASC questionnaires can not serve as a real cross-validation because the items sets of these questionnaires were not sampled independently. In fact, these items were specifically selected to be in accordance with the dimensional structure originally discovered with the APZ items.

Because Dittrich has adopted a methodological approach that was originally developed to discover broad dimensions of personality, his approach also rests – at least in part – on the "lexical hypothesis", which assumes that the most important variables in a domain are those used frequently in human communication, thus eventually becoming part of our natural language (Goldberg, 1981). Although the lexical hypothesis has been useful in the development of models of personality structure, it may have limitations with regard to research on ASCs because the western language may not be well equipped for the expression of rarely occurring ASCs, such as deep mystical experiences.

2.3.2 *Factorial Invariance*

2.3.2.1 *Similarity of correlation matrices*

Because sample sizes of the experimental studies were too small to test the hypotheses in each of the eleven stimulus conditions, the etiology-independency of items could only be evaluated in the following four groups of ASCs induction techniques: Hallucinogens of the first order (DMT, psilocybin, and THC; n = 82), hallucinogens of the second order (nitrous oxide; n = 38), sensory deprivation in a broader sense (perceptual deprivation, hypnagogic states, autogenic training, and hypnosis; n = 79), and sensory overload (stimuli of high variety; n = 60). Furthermore, because these groups were still too small to perform multivariate statistical analysis on each of them, the total data set (N =

259) had to be used for the dimensional analyses. Pooling the data over all stimulus conditions was justified on the grounds that similarity tests of the correlation matrices between all possible pairs of the four stimulus groups reached statistical significance. However, none of the 5 different similarity tests applied by Dittrich can formally test the hypothesis that the correlation matrices are equal. Hence, they provide little evidence that the correlation matrices from different groups could be reduced to exactly the same latent factor model.

Even if the equality of covariance matrices had been formally tested by a more stringent structural equation modeling approach (Jöreskog, 1971), it would not prove factorial invariance between groups because it is possible that the omnibus test of equality of covariance matrices indicates no group difference while hypothesis tests of the invariance of specific measurement or structural parameters of the latent factor model must be rejected (Byrne, 2009). Accordingly, many methodologists have concluded that the omnibus test of equal covariance matrices provides little guidance for testing the factorial invariance between groups (Brown, 2006). Instead, most authors recommend evaluating factorial invariance within a multiple groups CFA framework only. A major advantage of multiple groups CFA is that all potential aspects of invariance across groups can be examined by formal statistical hypothesis tests (Jöreskog, 1971). For instance, by setting increasingly restrictive constraints on the simultaneous parameter estimation of two CFA models, it can be assessed whether the number of factors and pattern of indicator-factor loadings is identical across groups (commonly referred to as configural invariance) and whether there is equality between groups with respect to factor loadings (weak factorial invariance), indicator intercepts (strong factorial invariance), and indicator residuals (strict factorial invariance; Meredith, 1993). Strict factorial invariance (cross-group equality in the loadings, intercepts, and residual variances) ensures that any systematic group differences in the means, variance or covariances for the measured variables are due to the common factors, rather than group differences in the factor structure. The establishment of strict factorial invariance is therefore a necessary condition for the correct interpretation of differences in the factor means (Brown, 2006; Millsap & Meredith, 2007). Unfortunately, neither the APZ nor the OAV and 5D-ASC questionnaires have ever been examined by multiple groups CFAs. Assessments of the factorial invariance of these questionnaires across different groups of ASC induction techniques and languages have entirely relied on descriptive measures of similarity derived from the comparison of EFA models.

2.3.2.2 *Item aggregates*

In the experimental studies, sample sizes were too small to compute EFAs on the whole set of 72 etiology-independent items in each of the four groups of stimulus conditions. Consequently, comparisons across stimulus groups had to be made on the basis of EFAs and cluster analyses computed on a set of nine item aggregates (so called item parcels) formed from the 49 variables having salient loadings in the EFA of the pooled data set. However, aggregating items to parcels is a highly dubious practice when the investigator's goal is to represent the dimensionality of the measurement space at the level of individual items (Little, Cunningham, Shahar, & Widaman, 2002). The use of item parcels can seriously distort the latent factor model when the item parcels are not unidimensional or when the uniquenesses (measurement error variances) of items

within a given parcel correlate with the unique or common factors of items in other parcels (Brown, 2006). Accordingly, numerous writers have suggested that only under conditions of unidimensionality should parceling be considered. Unfortunately, Dittrich has not assessed the dimensionalities of the items parcels for each stimulus conditions. Hence, the results of the EFA and cluster analysis comparisons of the four stimulus groups must be interpreted very cautiously.

2.3.2.3 *Tucker's coefficient of congruence*

Even if comparisons between stimulus groups had been made on the basis of EFAs and cluster analyses computed on the whole set of individual items, Dittrich's approach of investigating factorial invariance could be criticized for several other reasons. For instance, in order to assess the comparability of each latent dimension across different stimulus groups, Dittrich (1985, 1996) computed a similarity index commonly known as Tucker's coefficient of congruence (Burt, 1948; Tucker, 1951). Similar to the Pearson's product-moment correlation, the value of this statistic varies between +1 (perfect congruence) and -1 (perfect negative congruence), with zero indicating no linear association between the two factors. Its value is calculated by using only information from the two factor pattern (loading) matrices across the two groups. Although the coefficient of congruence has many desirable properties (see Lorenzo-Seva & ten Berge, 2006), it tends to overestimate similarity when the sign of all loadings is predominantly the same and the level of the loadings is approximately the same (Pinneau & Newhouse, 1964). In fact, the scale sensitivity problem of this similarity index is so severe that some researchers have discouraged it's usage entirely (e.g. Barrett, 2006).

The coefficient of congruence has also been criticized for being a mere descriptive measure of similarity for which no statistical estimation theory has been developed to test its significance (Zumbo, Sireci, & Hambleton, 2003). Although a solution to this problem has recently been proposed (Chan, Ho, Leung, Chan, & Yiu-Fai, 1999), it is still widely debated how large this coefficient should be to conclude that congruence holds to a reasonable approximation. In practice, the size of the coefficient of congruence is usually evaluated by applying some rule of thumb (e.g. see Lorenzo-Seva & ten Berge, 2006). However, it is unlikely that one decision rule would be appropriate for all situations because its size is affected by several variables. One such variable is the factor rotation of each solution under comparison. Because factor rotations are to a certain degree arbitrary, that is, two independent factor solutions may occupy markedly different orientations even though they are quite similar, Dittrich (1985, 1996) computed the coefficient of congruence after rotating one factor structure (the predictor matrix) to maximal agreement with the other (the criterion matrix) using orthogonal Procrustean rotation. Although widely applied in the old days of psychometric research and still recommended by some investigators for specific situations (McCrae, Zonderman, Bond, Costa, & Paunonen, 1996; Zumbo et al., 2003; Lorenzo-Seva & ten Berge, 2006), this procedure has many known weaknesses compared to the modern standard approach of measurement invariance evaluation, that is, multiple group confirmatory factor analysis (MG-CFA). One of the biggest problems is that Procrustean rotations have a tendency to capitalize on sample-specific subtleties, which – depending on the properties of the data being evaluated – may result in spuriously high and misleading coefficients of congruence (Nesselroade & Baltes, 1970; Korth & Tucker, 1975; Paunonen, 1997). By

using a Monte Carlo approach, Paunonen (1997) has demonstrated that capitalizing on chance and thus inflating the coefficients of congruence is dependent on the number of variables and the number of salient variables per factor (both having an inverse relation), the size of the salient variables' pattern coefficients (having a direct relation), and the number of factors (having a joint relation with the size of salients).

Unfortunately, Dittrich's comparisons of stimulus groups were made under conditions that increase the tendency to capitalize on chance the most, that is, a low number of variables, factors, and salient variables per factor and high loadings of salients. In fact, using the regression equation proposed by Paunonen (1997), it can be estimated that under Dittrich's conditions the upper 95% confidence limit of the random distribution of the coefficient of congruence is 0.89. Given that some coefficients of congruence reported by Dittrich (1985, 1996) were lower than that value, the evidence for an etiology-independent factor structure is actually quite weak. Furthermore, even if all reported coefficients of congruence had been higher than what might be expected by chance, it would only indicate statistically significant similarity. Significant similarity, however, does not imply factorial invariance. It only means that the factor patterns are not statistically independent. In other words, to demonstrate replicated factors in two samples, it is not sufficient to reject the null hypothesis of zero factor congruence. Instead, one has to show that the coefficient of congruence is not statistically different from one (Paunonen, 1997). In order to test the latter hypothesis, Chan et al. (1999) have proposed a bootstrap-procedure. Using this procedure in a Monte Carlo study, they estimated the minimum sizes of the coefficient of congruence that were needed to not reject the null hypotheses of equal factor loadings under eight different model conditions. Interestingly, the average critical values determined in that study are generally much higher than what commonly applied rules of thumb suggest and thus further call into question that the congruence coefficients reported by Dittrich (1985, 1996) are high enough to demonstrate factorial invariance. Moreover, even if all coefficients of congruence indicated equality of factor loadings, it would still not satisfy the criteria of measurement invariance. As mentioned above, in order to establish measurement invariance between groups, it is also necessary to show that the models under comparison are invariant with respect to unique variances and indicator intercepts. The only possible way of addressing these issues is by carrying out CFAs involving the analysis of mean in addition to covariance structures (Millsap & Meredith, 2007).

2.3.2.4 *Cohen's Kappa*

Previous studies examining the replicability of the factorial structures of the APZ, OAV, and 5D-ASC questionnaires in different populations (Dittrich, von Arx, & Staub, 1985; Dittrich, 1985; Bodmer, 1989; Bodmer et al., 1994; Dittrich, 1994; Braun, 1997; Bodmer, 1999) have also relied on Cohen's $\varkappa$ (Cohen, 1960) as a descriptive measure of latent structure similarity. The coefficient $\varkappa$ estimates the similarity of two factors from different solutions by comparing the patterns of salient and non-salient loadings (Guadagnoli & Velicer, 1991). It has also been used to compare two factorial solutions as a whole by comparing the distributions of the highest loadings of each item. The latter approach has been used exclusively in the existing literature on the APZ, OAV, and 5D-ASC questionnaires. Similar to the coefficient of congruence, the computation

of $\varkappa$ is entirely based on information of the two factor loading matrices across the two groups. However, because the factor loadings are treated as binary values, that is, they are either salient or not, some information is lost in the computation of the index. Furthermore, the size of $\varkappa$ is dependent on an arbitrarily defined cut-off value above which variables are determined to be salient. Because $\varkappa$ does only reflect differences in the distribution of salient items across two groups, it is merely a descriptive measure of configural invariance, which is the weakest form of factorial invariance. As mentioned above, configural invariance is a necessary but not sufficient condition for measurement invariance. Furthermore, $\varkappa$ seems to be a very lenient indicator of configural invariance. For instance, even if all 18 VRS items of the OAV questionnaire were wrongly assigned to the OBN factor in one sample, an overall comparison of that factorial solution with the hypothetically correct solution would result in a $\varkappa$ coefficient of 0.56, which according to the guidelines of Landis and Koch (1977) would still be interpreted as indicating fair to good agreement. Obviously, testing the factorial invariance across two samples by the $\varkappa$ coefficient is so insensitive to large differences in the loading patterns that a falsification of the hypothesis of similar loading matrices is almost impossible to achieve. Stronger falsification testing would have been possible by computing $\varkappa$ for each factor. However, this has not been attempted in any previous APZ, OAV, and 5D-ASC questionnaire validation study.

2.3.3 *Binary Items*

Another methodological shortcoming of Dittrich's dimensional analyses is that binary items were treated as if they were continuous. This potentially biased not only the dimensional analyses of the APZ questionnaire, which has a dichotomous response format, but also the validation of the 5D-ASC, for which a set of dichotomized BETA and APZ items were analyzed (see Braun, 1997). The main problem is that Dittrich and his colleagues estimated associations between binary items with the Pearson correlation coefficient, which in the case of binary variables is called φ coefficient. It has long been known that the φ coefficient can be severely attenuated if the items differ markedly by their difficulties (J. B. Carroll, 1945). In fact, the correlation between two binary items can reach 1.0 only when the items have exactly the same endorsement probabilities. Factor analyses on the basis of φ coefficients therefore usually lead to an underestimation of factor loadings (Woods, 2002). Furthermore, because items cluster together according to threshold, in addition to content, "pseudofactors" that are artifacts of item difficulty or extremeness can emerge (Reise, Waller, & Comrey, 2000; Kubinger, 2003). Indeed, an analysis of variance applied to the item difficulties obtained from the total sample of Dittrich's experimental studies reveals highly significant differences between the primary three dimensions ($F(2, 46) = 6.48$, $p = 0.003$). It is therefore quite likely that the dimensional analysis of the APZ has been distorted to some degree by the unequal item difficulties.

Future studies analyzing the factorial structure of the APZ should be based on tetrachoric correlations. Unlike Pearson correlations, tetrachoric correlations of binary items are not attenuated if the items differ by their difficulties (J. B. Carroll, 1945). Hence, factor analyses based on tetrachoric correlations are much better suited to reveal the underlying factorial structure of a set of binary items (Kubinger, 2003; Reise et al., 2000).

Although factor analyzing tetrachoric correlations would be a big improvement, it can still lead to vastly inflated chi-square values of model fit and underestimated standard errors of estimates when the factor model is estimated by the ordinary maximum likelihood estimator (B. O. Muthén, 1989). In recent years, several estimators have been developed that are better suited for analyzing categorical items; for example, weighted least squares (WLS), weighted least squares mean and variance adjusted (WLSMV), and unweighted least squares (ULS). Simulation studies (e.g., Flora & Curran, 2004) have shown that the WLSMV estimator gives the most accurate results, especially with small to moderate sample sizes. It is therefore strongly recommended to use this estimator in future factor analyses of the APZ.

2.3.4 *Reliability*

In all APZ, OAV, and 5D-ASC questionnaire validation studies, Dittrich and his colleagues have relied on Cronbach's α to demonstrate scale reliabilities. However, the use of Cronbach's α as a reliability estimate is problematic because it rests on two statistical assumptions that are rarely met in practice, namely, unidimensionality and essential tau equivalence (Revelle & Zinbarg, 2009). Unidimensionality means that the common variance between the items of a scale can be totally explained by one common general factor and that the scale contains no group factors. Essential tau-equivalence means that all items of a given scale are equally strongly related to one common factor. Within the CFA framework, unidimensionality is demonstrated when there are no residual correlations and essential tau-equivalence is demonstrated when all factor loadings are equal (Brown, 2006). If only the condition of essential tau-equivalence is violated, α is a conservative estimate of scale reliability, that is, it will underestimate scale reliability. However, if the condition of unidimensionality is violated – and this is very often the case, especially when the scale contains many items – α can both under- and overestimate scale reliabilities, depending on the underlying measurement parameters (Zinbarg, Revelle, Yovel, & Li, 2005; Brown, 2006). The use of Cronbach's α has therefore long been criticized by several authors (Revelle, 1979; Huysamen, 2006; Sijtsma, 2009; Green & Yang, 2009). Recently, criticism has become more widespread because many alternative reliability estimates are now available that are less biased that can be more easily interpreted (Raykov, 2001; Zinbarg, Revelle, Yovel, & Li, 2005; Revelle & Zinbarg, 2009).

Because Dittrich and his colleagues have neither assessed essential tau equivalence nor unidimensionality before calculating Cronbach's α, the reliabilities of the APZ, OAV, and 5D-ASC scales that have been reported so far must be taken with a grain of salt. Future studies should either test the necessary assumptions before calculating Cronbach's α or assess the reliabilities of the scales by using newly developed reliability estimates that make less strong assumptions and that can be more easily interpreted. Among the most promising reliability estimates (at least for continuous items) to date are McDonalds ω_h and ω_{tot} (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005; Revelle & Zinbarg, 2009). Whereas ω_h estimates the amount of variance in a scale attributable to one common factor, also referred to as general factor saturation, ω_{tot} estimates the amount of variance due to all common factors (i.e., group factors and general factor; Zinbarg, Revelle, Yovel, & Li, 2005).

In addition to the use of Cronbach's α, the way Dittrich and his colleagues have assessed retest-reliabilities (i.e. simple correlations of summated scores) is also suboptimal. Test-retest correlations of summated scores have been shown to be positively biased by the failure to take into account correlated uniquenesses and negatively biased by the failure to control measurement error (Marsh & Hau, 2007). A much better approach is to estimate test-retest correlations within the so called correlated uniquenesses model (e.g., see Brown, 2006).

2.4 DISCUSSION AND CONCLUSIONS

This chapter has critically reviewed the available literature on the development and validation of Dittrich's APZ, OAV, and 5D-ASC questionnaires from a modern methodological perspective. It has demonstrated that the existing studies have many methodological shortcomings and that some of them seriously weaken the conclusions drawn by Dittrich and his colleagues.

To be fair, it should be noted that most of the reviewed studies are relatively old. Specifically, they were mostly conducted between the early eighties and the mid nineties of the last century, when many of the statistical methods that I have recommended in this chapter were not yet implemented in statistical packages. In fact, even today some of these methods are not widely available. For instance, the calculation of polychoric correlations are still not supported by the newest version of SPSS, and the WLSMV estimator is only available in the statistical software M*plus* (L. K. Muthén & B. O. Muthén, 2007), although it should be soon implemented in lavaan (Rosseel, Oberski, & Byrnes, 2011), which is a freely available add-on package to R (R Development Core Team, 2011). Some of the weaknesses highlighted in this chapter must also be reconciled with economic constraints. Conducting experiments in which ASCs are induced by various means is costly because subjects must be carefully selected and monitored. Thus, it would have been quite difficult to test Dittrich's hypotheses with larger samples, more stimulus conditions, and other item sets.

Nevertheless, many of the methodological problems mentioned in this chapter have been described in the literature long before Dittrich (1985) conducted his first investigations on the structure of ASCs. Furthermore, the theoretical foundations for structural equation modeling, CFA, and MG-CFA, which now have become standard methods of psychometric investigations and which are associated with many of the methodological and statistical advances in quantitative psychology in the last two decades (Marsh, Lüdtke, et al., 2010), were mostly established in the late nineteen-sixties and seventies (e.g., Jöreskog, 1969, 1971), and the first statistical software that could fit such models (i.e. Lisrel) also became available at that time. Of course, when Dittrich (1985) wrote his "opus magnum" in the early eighties, such methods were still quite experimental, but they were already quite established in the mid nineties, when the OAV and 5D-ASC questionnaires were developed and/or validated. Thus, the inappropriate use of exploratory methods (EFA) for hypothesis testing and the omission of confirmatory methods (CFA and MG-CFA) – especially in the newer studies – cannot solely be excused for historical reasons.

As a proponent of critical rationalism (Popper, 1959), Dittrich has repeatedly exposed his hypotheses to refutation through falsification tests. Since none of his falsification

tests have proved his hypotheses wrong, he concluded that his hypotheses have survived falsification. Thus, even though, as a critical rationalist, he would never claim that this hypotheses were true, he has concluded that there is considerable evidence that they are not wrong. However, as I have shown above, Dittrich's criteria for falsification were often so lenient that a refutation of his hypotheses was almost impossible to achieve. The most striking example is the use of Cohen's $\varkappa$, which indicates fair to good agreement between two factor structures even though the loading matrices strongly differ.

Thus, I consider it an exaggeration to claim that the hypotheses have survived considerable falsification testing. Similarly, I regard it as premature to call the OBN, DED, and VRS constructs "etiology-independent" dimensions of ASCs. Although it is still possible that Dittrich's hypotheses are true, stronger falsification tests are clearly needed to substantiate them. Methods of the structural equation modeling framework, such as CFA, MG-CFA and ESEM (e.g., see Marsh, B. O. Muthén, et al., 2009), are currently the most appropriate methods for testing Dittrich's original hypotheses and have the potential to overcome many of the methodological shortcomings described in this chapter. Studies that apply these methods to Dittrich's ASC questionnaires are therefore highly warranted.

Part II

EMPIRICAL STUDIES

3

ACUTE, SUBACUTE AND LONG-TERM SUBJECTIVE EFFECTS OF PSILOCYBIN IN HEALTHY HUMANS: A POOLED ANALYSIS OF EXPERIMENTAL STUDIES

3.1 ABSTRACT

Psilocybin and related hallucinogenic compounds are increasingly used in human research. However, due to limited information about potential subjective side-effects, the controlled medical use of these compounds has remained controversial. We therefore analyzed acute, subacute, and long-term subjective effects of psilocybin in healthy humans by pooling raw data from eight double-blind placebo-controlled experimental studies conducted between 1999 and 2008. The analysis included 110 healthy subjects who had received 1-4 oral doses of psilocybin (45-315 μg/kg body weight). Although psilocybin dose-dependently induced profound changes in mood, perception, thought, and self-experience; most subjects described the experience as pleasurable, enriching, and non-threatening. Acute adverse drug reactions, characterized by strong dysphoria and/or anxiety/panic, occurred only in the two highest dose-conditions in a relatively small proportion of subjects. All acute adverse drug reactions were successfully managed by providing interpersonal support and did not need psychopharmacological intervention. Follow-up questionnaires indicated no subsequent drug abuse, persisting perception disorders, prolonged psychosis, or other long-term impairment of functioning in any of our subjects. The results suggest that the administration of moderate doses of psilocybin to healthy, high-functioning, and well-prepared subjects in context of a carefully monitored research environment is associated with an acceptable level of risk.

3.2 INTRODUCTION

Psilocybin (4-phosphoryloxy-*N*,*N*-dimethyltryptamine) is an indoleamine or serotonin-like hallucinogen and the main psychoactive principle of a group of hallucinogenic fungi of the genus Psilocybe, also often referred to as "magic mushrooms" (Hofmann, 1968). Psilocybe mushrooms occur throughout the world and their human use in medical and religious rituals dates back for centuries, if not millennia (Stamets, 1996; Guzmán et al., 2000).

Modern psychopharmacological research with psilocybin begun with the discovery of the cultic use of Psilocybe mushrooms by Mesoamerican Mazatec Indians in 1955 (Wasson, 1958). Psilocybin and psilocin were identified at Sandoz Laboratories as the psychoactive compounds of Psilocybe mushrooms and synthesized by the renowned Swiss chemist Albert Hofmann (1958), who some 15 years earlier also discovered the chemically related ergoline hallucinogen LSD. Soon after, synthetic psilocybin was marketed by Sandoz under the name Indocybin® for basic psychopharmacological and therapeutic clinical research (Hofmann, Heim, Brack, Kobel, et al., 1959; Passie, 1995; Passie, Seifert, et al., 2002).

Early clinical studies in the 1960s and 1970s demonstrated that psilocybin produces an altered state of consciousness (ASC) similar to LSD that is characterized by marked alterations in perception, mood, and thought, including changes in the experience of time, space, and self, that are rarely experienced otherwise except in dreams, religious exaltation, and acute psychoses (Isbell, 1959; Rümmele & Gnirss, 1961; Leuner, 1962; Wolbach et al., 1962; Fischer, 1971; Geyer & Vollenweider, 2008). In these states, perceptual hypersensitivity, illusions, and pseudohallucinations (i.e., hallucinations with intact reality testing and insight) are common (Leuner, 1962; Hill, Fischer, & Warshay, 1969; Fischer, Hill, & Warshay, 1969; Fischer, Hill, Thatcher, & Scheib, 1970).

Intensification of affective responses, enhanced ability for introspection, regression to primitive and childlike thinking, and activation of vivid memory traces with pronounced emotional undertones can also occur (Leuner, 1971). Psychophysiological and pharmacological studies revealed that psilocybin had a much shorter duration of action than LSD (4-6 h instead of 8-12 h; Cerletti, 1959). Although - apart from the duration of action - the effects of both drugs were found to be highly similar in a controlled study (Hollister & Hartman, 1962), clinical observations indicated that psilocybin tended to produce less anxiety, panic reactions, affective disturbances, and milder vegetative side-effects than LSD (Heimann, 1962; Nieto, 1962; A. E. David & J. M. David, 1961; Clark, 1968; Leuner, 1968; Passie, 1995). Hence, many hallucinogen researchers valued psilocybin as a useful substitute for the earlier discovered LSD to explore the neural basis of ASCs including the basis of hallucinations, religious, spiritual, and psychotic dimension, while for others it was an attractive adjunct in psychodynamic-oriented psychotherapy to bring the unconscious into conscious (Leuner, 1971; Nichols, 2004; Nichols & Chemel, 2006).

Throughout the 1960s, LSD and related drugs became increasingly associated with cultural rebellion, were widely popularized as drugs of abuse, and depicted as dangerous. Consequently, around 1970, LSD and related drugs were scheduled in the most restrictive category in most countries. Accordingly, human research on psychedelics became severely restricted, funding became difficult, and interests in therapeutic use of these drugs faded.

The 1990s witnessed a re-emergence of human hallucinogen research in Europe, particularly with the development of new brain imaging techniques, sophisticated neuropsychological approaches, and neuropharmacological findings that have supported hallucinogens as models of at least some aspects of natural occurring psychosis (for a review, see Geyer & Vollenweider, 2008). Many of these studies conducted in our and other laboratories used psilocybin as a tool to investigate the neural underpinnings of psychotic symptom formation including ego-disorders and hallucinations (Vollenweider, 1992; Vollenweider, Leenders, Scharfetter, P. Maguire, et al., 1997; Vollenweider, Vollenweider-Scherpenhuyzen, et al., 1998; Gouzoulis-Mayfrank, Schreckenberger, et al., 1999; Gouzoulis-Mayfrank, Thelen, et al., 1999; Vollenweider & Geyer, 2001; Hasler, Grimberg, et al., 2004) or to explore the effect of psilocybin on cognitive and visual processes (Spitzer, Thimm, et al., 1996; Umbricht, Koller, et al., 2002; Umbricht, Vollenweider, et al., 2003; Carter, Pettigrew, Burr, et al., 2004; Carter, Burr, et al., 2005; Carter, Pettigrew, Hasler, et al., 2005; Carter, Hasler, et al., 2007; Kometer et al., 2011), time perception (Wackermann et al., 2008; Wittmann et al., 2007), and on sensory gating and

its relation to cognitive alterations (Gouzoulis-Mayfrank, Heekeren, Thelen, et al., 1998; Vollenweider, Csomor, et al., 2007).

Recent work has also explored the acute and long-term subjective effects of psilocybin in hallucinogen-naïve healthy subjects and found that 14 months after the experiments about two-third of participants still rated the experience as among the most personally meaningful and spiritually significant of their lives (Griffiths, Richards, McCann, & Jesse, 2006; Griffiths, Richards, Johnson, et al., 2008). In addition, several recent studies have re-investigated the tolerability and efficacy of psilocybin in the treatment of anxiety related advanced-stage cancer (Grob et al., 2011) and obsessive-compulsive disorders (Moreno et al., 2006). Other studies have focused on the pharmacokinetics (Hasler, Bourquin, Brenneisen, Bär, & Vollenweider, 1997), metabolism (Hasler, Bourquin, Brenneisen, & Vollenweider, 2002), dose-dependent effects (Hasler, Grimberg, et al., 2004), and receptor mechanism of psilocybin (Ametamey et al., 1998; Vollenweider, Vollenweider-Scherpenhuyzen, et al., 1998; Vollenweider, Vontobel, Hell, & Leenders, 1999; Vollenweider, Hasler, & Kometer, 2008; Hasler, Quednow, et al., 2009). For example, we have shown that the selective 5-HT_{2A} receptor antagonist ketanserin blocks the hallucinogenic effects of psilocybin in human subjects (Vollenweider, Vollenweider-Scherpenhuyzen, et al., 1998), providing strong evidence for the link between 5-HT $_{2A}$ receptor activation and hallucinosis (Sanders-Bush, Burris, & Knoth, 1988). Moreover, a recent animal study found a novel mechanism of functional interaction between 5-HT_{2A} and mGluR2 receptors and suggests that specific mGluR2 agonists may block the effect of hallucinogens such as psilocybin in humans (González-Maeso, Weisstaub, et al., 2007; González-Maeso, Ang, et al., 2008; González-Maeso & S. C. Sealfon, 2009). Although molecular and mechanistic studies in animal are pertinent, translational research in human subjects and particularly the establishing of the links between the mechanism of action of hallucinogens and the subjective effects in humans is essential (Vollenweider, 2001).

Although serotonergic hallucinogens such as psilocybin are considered relatively safe physiologically and do not produce dependence (Leuner, 1981; Nichols, 2004; Johnson, Richards, & Griffiths, 2008), there is limited information on the acute tolerability and potential long-term psychological effects of psilocybin. Moreover, given that many of the early human studies with psilocybin were poorly standardized and lacked adequate control groups or follow-up measures, or often had small and unrepresentative sample sizes, it is difficult to draw inferences with this work, e.g., on dose-response effects and the incidence of acute and subacute distress and side effects.

The purpose of this paper is to provide further information about the acute, subacute, and potential long-term subjective effects of psilocybin administration in healthy human subjects in a controlled experimental setting. The present analysis is based on data of 8 double-blind placebo-controlled psilocybin studies that were conducted in our laboratory during the past 10 years. The data report on acute, subacute, and long-term effects of 227 individual psilocybin sessions obtained in 110 subjects using validated instruments to assess various aspects of consciousness, mood, psychological and physical side effects. Whereas parts of the data on the acute effects of psilocybin were previously published (see Table 1), the data on subacute and long-term effects/side-effects have not been presented elsewhere before.

3.3 METHODS

3.3.1 *Study Description*

Data from eight experimental studies involving psilocybin administration to healthy human subjects carried out between 1999 and 2008 at our research facility were pooled for the present analysis (Table 1). Earlier psilocybin studies were not included because no long-term follow-up measurements were obtained in those experiments. All eight studies were approved by the Ethics Committee of the University Hospital of Psychiatry, Zürich, and the use of psilocybin was authorized by the Swiss Federal Office of Public Health, Department of Pharmacology and Narcotics, Bern. In each study, a double-blind placebo-controlled within-subject design was used. Depending on the study, subjects were tested on 2-5 experimental days, each separated by at least 14 days to avoid carry-over effects. Each volunteer received placebo and 1-4 different oral doses of psilocybin in a randomized and counterbalanced order. Additionally, in one study subjects also received pre-treatments with the 5-HT_{2A} antagonist ketanserin and placebo. Psilocybin doses ranged from 45 µg/kg to 315 µg/kg body weight (absolute doses: 2-28 mg). For a detailed description of the administered psilocybin doses in each study, see Table 1.

Experimental sessions of studies 3 and 7 were conducted at the positron emission tomography center of the University Hospital, Zürich, while all other study sessions took place at the University Hospital of Psychiatry, Zürich. All subjects were instructed to have a light breakfast prior to the experiments. Before testing began, blood pressure and heart rate were measured and subsequently monitored at hourly intervals throughout the day. Subjects finished participation of the study approximately 7 h after psilocybin administration and were examined by the principal investigator before being deemed fit for release. Subjects were asked not to engage in demanding work after the psilocybin session and to contact the research staff if any adverse events occurred. These procedures are generally similar to those subsequently used and recommended by Johnson, Richards, and Griffiths (2008).

3.3.2 *Subjects*

In all studies, psilocybin (4-phosphoryloxy-*N*,*N*-dimethyltryptamine) was obtained through the Swiss Federal Office for Public Health, Bern. Psilocybin capsules (1 mg and 5 mg) were prepared at the Pharmacy of the Cantonal Hospital of Aarau, Switzerland. The psilocybin and lactose placebo were prepared in gelatin capsules of identical appearance.

3.3.3 *Psychometric Ratings of Acute and Post-Acute Effects*

All eight studies included in the present analysis used the Five Dimensions of Altered States of Consciousness rating scale (5D-ASC; Braun, 1997; Dittrich, Lamparter, & Maurer, 2006) and six studies also used the Eigenschaftswörterliste 60-S (EWL-60-S; Janke & Debus, 1978, 1986) to assess acute and subacute subjective drug effects.

The 5D-ASC questionnaire is a psychometrically improved and extended version of the original APZ questionnaire (Dittrich, 1998). The 5D-ASC is a visual analogue

Table 1: Pooled Psilocybin Studies

Study description	Psilocybin dose condition	Subjects receiving at least one dose	Number of administered psilocybin doses				Publication
			Very low dose 45 µg/kg	Low dose 115-125 µg/kg	Medium dose 215-260 µg/kg	High dose 315 µg/kg	
1) Dose-effect study on acute psychological and physiological effects of psilocybin	1) 45 µg/kg 2) 115 µg/kg 3) 215 µg/kg 4) 315 µg/kg	8	8	8	8	8	Hasler et al. 2004
2) Acute effects of psilocybin on cognitive functions and subjective experience.	1) 115 µg/kg 2) 215 µg/kg 3) 315 µg/kg	16		16	16	16	In progress
3) Effects of psilocybin on brain activity using H_2O-PET.	260 µg/kg	12			12		In progress
4) Effects of psilocybin on prepulse inhibition of startle in healthy human volunteers.	1) 115 µg/kg 2) 215 µg/kg 3) 315 µg/kg	20		17	17	18	Vollenweider et al. 2007
5) Effects of psilocybin on the rate and rhythmicity of perceptual rivalry alternations.	1) 115 µg/kg 2) 250 µg/kg	12		12	12		Carter et al. 2004 & 2005 Wittmann et al. 2007 Wackermann et al. 2008
6) Investigation on the relationship between attention, working memory, and the serotonin 1A and 2A receptors using psilocybin and ketanserin.	1) 215 µg/kg 2) 215 µg/kg after ketanserin pretreatment	10			10		Carter et al. 2005 Carter et al. 2007
7) Effects of psilocybin on visual processing: An EEG study.	1) 125 µg/kg 2) 250 µg/kg	21		21	18		Kometer et al. 2011
8) Serotonin 5-HT_{2A}- receptor dynamics in the human brain following psilocybin stimulation	250 µg/kg	11			11		In progress
Total		110	8	74	104	42	

self-rating scale consisting of 94 items, assessing five primary dimensions and one global dimension of ASCs. The primary dimensions are comprised of several item clusters and can be described as follows: (1) Oceanic Boundlessness (OBN) measures derealization and depersonalization accompanied by changes in affect ranging from heightened mood to euphoria and/or exaltation, and alterations in the sense of time. The corresponding item clusters are *positively experienced derealization, positive experienced depersonalization, changed sense of time, positive mood, and mania like experience*. (2) Dread of Ego Dissolution (DED) measures ego-disintegration associated with loss of self-control, thought disorder, arousal, and anxiety. The item clusters are *negatively experienced derealization, thought disorder, paranoia, loss of thought control, and loss of body control*. (3) Visionary Restructuralization (VRS) measures alterations in perception and meaning. The Item clusters are *elementary hallucinations and illusions, scenery hallucinations, synesthesiae, changed meaning of percepts, facilitated recollection, and facilitated imagination*. (4) Auditory Alterations (AUA) measures auditory illusions and auditory (pseudo-) hallucinations. (5) Vigilance Reduction (VIR) relates to states of drowsiness, reduced alertness, and impaired cognitive function. The OBN, DED, and VRS dimensions have been shown to be common to a range of altered states of waking consciousness of varying aetiology and intensity, while the AUA and VIR dimensions are hypothesized to occur only under certain stimulus conditions (Dittrich, von Arx, & Staub, 1985; Dittrich, 1998). Items from the OBN, DED, and VRS dimensions are therefore summed to the Global Altered States of Consciouness (G-ASC) scale. 5D-ASC administration time in the pooled data set varied between studies from 60 to 270 min after drug intake. In all studies, subjects were instructed to rate their whole experience by the 5D-ASC retrospectively from the moment of drug intake to the respective measuring time point.

The EWL-60-S is a self-rating scale that was developed for the multidimensional assessment of mood states and condition (Janke & Debus, 1986). The EWL-60-S consists of a list of 60 adjectives (e.g., 'anxious', 'tired', 'sociable') that are rated by subjects as to how well they describe their present state. Subjects must choose from four possible answers ('not at all', 'somewhat', 'quite', or 'strongly'). EWL-60-S items can be broken down into 15 mood states and conditions: "Efficiency-activation", "Concentration", "Inactivation", "Tiredness", "Drowsiness", "Extroversion", "Introversion", "Self-confidence", "Heightened mood", "Emotional excitation", "Sensitivity", "Aggression-anger", "Apprehension-anxiety", "Depressiveness", and "Dreaminess". Depending on the study, the EWL-60-S was administered at one to four time points during the course of an experimental session and between 60 min and 24 h after drug administration (see statistical analysis section for more details on the EWL-60-S administration time points in each study). At each time of assessment, subjects were instructed to rate their present state.

3.3.4 *Psychometric Rating of Subacute Side-Effects*

Subacute side effects were assessed in six studies (study # 1, 2, 4, 5, 6, 8) by the List of Complaints (LC; von Zerssen, 1971). This self-rating scale consists of a list of 65 common somatic and psychological ailments, which can be summed to a global score of general discomfort. Subjects were asked to rate whether each symptom is present or not at the time of assessment. In all 6 studies, the LC was administered 24 h after drug intake.

3.3.5 *Long-Term Follow-Up*

Long-term psilocybin effects were assessed by an investigator-constructed follow-up questionnaire that covered the following areas of content:

1. Ratings of acute drug experiences in retrospect:

Subjects were asked the following question: "How do you rate the acute drug experience during the experiment in retrospect?" For each of six adjectives (pleasant, enriching, frightening, unpleasant, influential, and nothing special), subjects had to choose one of three possible answers (very much, medium, or not).

2. Changes in values and attitudes:

Subjects were asked the following questions: Did the experiment with psilocybin cause changes in (a) world view, (b) values, (c) awareness of personal problems, (d) the relationship to your body, (e) relationships to other people, (f) professional relationships, (g) the relationship to the environment/nature, (h) aesthetic experiencing, and (i) in the attitude to ASCs? For each item, subjects had to choose from three possible answers (positive change, negative change, or no change)

3. Changes in drug consumption habits:

Subjects were asked whether they had changed their consumption habits of any psychoactive drug after the experiments. For each drug that was consumed either more or less often than before, subjects were asked to give further details on frequency of use, dosages, route of administration, and setting of use. Subjects were also asked whether they considered the described changes as a consequence of their drug experience during the experiments.

4. Spontaneously occurring ASCs before and after the experiments and flashbacks:

Subjects were asked to describe frequencies, durations, circumstances, and symptoms of ASCs that spontaneously occurred before and/or after the experiments and whether they interpreted these ASCs as a flashback-like re-experiencing of acute drug effects.

5. Negative changes in psychological well-being and/or mental functions:

Subjects were asked to report the intensity, duration, and frequency of any experienced negative change in well-being and/or mental functions after the experiments. Sleeping, memory, and concentration problems, as well as mood swings, anxiety, and reactivation of old problems were directly listed in the questionnaire, but further symptoms could be described by the subjects if necessary. Long-term follow-up questionnaires were mailed to the study subjects (including drop-outs) 8-16 months after completion of their last experimental session. All follow-up questionnaires were obtained after subjects had been paid and therefore were more likely to report adverse effects.

3.3.6 *Statistical Analysis*

All statistical analysis were performed using the freely available statistical package R© (Version 2.8.1; R Development Core Team, 2008). Since all analyzed studies used within-subject designs and since there is considerable heterogeneity between studies (e.g., differences in setting, study manager, and experimental procedure), our pooled data of acute and subacute psilocybin effects is structured hierarchically with repeated measurements nested within subjects and subjects nested within studies. Because drug dose conditions only partially overlap between studies and some subjects prematurely

dropped out, our data set is also unbalanced with respect to drug dose condition. To account for observational heterogeneity and the lack of balance, we used mixed-effects models, which readily handle unbalanced and missing data and allow all observational units to contribute information to the analysis (Pinheiro & Bates, 2000). We used the R add-on package nlme (Pinheiro, Bates, DebRoy, Sarkar, & the R Core team, 2008) to fit mixed effects models. To minimize a potential bias arising from drop-outs, all available data of drop-outs were included in the statistical analyses on acute, subacute, and long-term psilocybin effects.

To investigate the acute effects of drug dose on the five ASC dimensions (OBN, DED, VRS, AUA, and VRS), 5D-ASC data assessing peak drug effects were pooled over all eight studies. In studies where the 5D-ASC questionnaire was administered more than once during an experimental session, data from the measuring time points yielding the highest mean total score were used. The marginally differing psilocybin dose conditions 115 and 125 µg/kg, as well as 250 and 260 µg/kg were combined because t-tests between these pairs of dose conditions did not reveal significant differences of subjective drug effects measured by the global and primary dimensions of 5D-ASC. The pooled 5D-ASC data were analyzed by linear mixed effects models with treatment (placebo, 45, 115-125, 215, 250-260, and 315 µg/kg) as a fixed effects factor and study and subjects within studies as random effects factors. Random effects were modeled as random intercepts without random slopes. Akaike's information criterion (AIC) values were used to decide on appropriate correlation structures of random effects in model specifications. In cases of variance heteroscedasticity, we used a weighting procedure to correct for unequal variances between groups. Statistical assumptions were checked graphically by plotting residuals against predicted values and by normal quantile-quantile plots of residuals and random effects. In each fitted mixed-effects model, the shape of the psilocybin dose-response relationship was evaluated by means of orthogonal polynomial contrasts. Additionally, when significant treatment effects were detected, one-tailed Dunnett contrasts were applied using the R add-on package multcomp (Hothorn, Bretz, & Westfall, 2008) in order to find the minimum effective dose. To determine the proportions of subjects experiencing strong subjective drug effects cumulative distributions of the three aetiology-independent dimensions (OBN, DED, and VRS) were inspected for each dose condition. Scale values above 70 percent of the maximum possible score were considered as strong subjective drug effects.

The analysis of EWL-60-S data was restricted to the placebo and medium dose psilocybin conditions because these were the only drug conditions that occurred in all six studies using the EWL-60-S. For the analysis of time dependent effects of psilocybin, the EWL-60-S administration time, which widely varied across studies, was categorized as follows: t_1 = 60-95 min, t_2 = 160-180 min, t_3 = 260-400 min, and t_4 = 24 h after drug intake. Five studies (study # 1, 4, 5, 6, and 8) had used the EWL-60-S during t_1 and t_3. Three studies (study # 3, 4, and 8) administered the EWL-60-S during t_2 and three studies (study # 1, 5, and 8) used the EWL-60-S at t_4. The EWL-60-S subscales were analyzed by mixed effects models using the fixed effects factors treatment (placebo vs. 215-260 µg psilocybin) and time (t_1, t_2, t_3, and t_4) and the hierarchically nested random effects factors study, subject, and subject within treatment. For the assessment of acute effects of psilocybin, which should occur at t_1 and t_2 and to a lesser degree at t_3 and t_4, the interaction of Drug $\times$ Time was considered as the main source of information.

Longer-lasting psilocybin effects were determined by significant main-effects of drug in the absence of Drug × Time interactions. *P*-values were adjusted by Holm's method in order to maintain a type I error rate of $p < 0.05$ over all 30 hypothesis tests arising from the evaluations of the main effect of drug and the Drug × Time interaction in each of the 15 subscales.

Subacute side effects measured by the LC were analyzed on the total scale as well as on the item level. Scores of the total scale were square-root transformed to reduce positive skew and subsequently analyzed by a linear mixed effects model with the within-subject factor treatment (placebo, 115-125 , 215, 250, and 315 µg/kg) and the nested random effects factors study and subject. Differences between all possible pairs of drug dose conditions were assessed by specifying Tukey's contrasts and adjusting *p*-values by Holm's method. On the item level, frequency differences of single complaints between three different doses of psilocybin (125, 215, and 315 µg/kg) were analyzed by Cochran Q tests using data from studies # 1, 2, and 4. In order to analyze single complaints in the largest possible sample, we also used McNemar tests to compare items frequencies on placebo and medium dose psilocybin (215-250 µg/kg) taking data from 6 studies (study # 1, 2, 4, 5, 6, and 8).

Because all participants had received psilocybin and no control group was available for comparison on responses to the long-term effects questionnaire, responses were analyzed with descriptive statistics only. Data of forced-choice items were analyzed by calculating absolute numbers and proportions of responses, whereas data of free-response items were either categorized to calculate sum and percentage scores or summarized in the text.

3.4 RESULTS

3.4.1 *Sample Characteristics and Drop-Outs*

In the studies carried out at our research facility between 1999 and 2008, 227 experimental sessions involving psilocybin administration were conducted. In total, 8 very low doses (45 µg/kg body weight), 74 low doses (115-125 µg/kg), 104 medium doses (215-260 µg/kg), and 41 high doses (315 µg/kg) of psilocybin were administered. The number of subjects who received at least one dose of active psilocybin was 110 (59 males and 51 females). Subjects were between the age of 20 and 47 (M ± SD: 26.9 ± 5.5 years) and exclusively Caucasians. 56% of subjects were university students and 33% were university graduates. 60% of subjects had no prior experience with a classical hallucinogen (LSD, psilocybin, DMT, or mescaline); 20% had consumed it 1-10 times in a lifetime; and 20% had consumed it more than 10 times in a lifetime but maximally 6 times per year. 90% of subjects had smoked cannabis at least once in a lifetime.

Of the 110 subjects included in the pooled analysis, seven subjects had prematurely dropped out after having received at least one dose of active psilocybin. A review of the study protocols and questionnaire data of these subjects revealed that in two cases, drop-outs were due to technical reasons and unrelated to drug effects (one subject moved to another country and another subject had too many EEG-artifacts). In the remaining five cases, two subjects had an unusually intense reaction to a low dose of psilocybin and were therefore excluded by the study manager due to safety

considerations. Another subject experienced a transient hypotonic reaction (systolic and diastolic blood pressure: 86/63 mm/Hg) with dizziness, fainting, and vomiting after having received 115 µg/kg of psilocybin and was therefore also excluded from further psilocybin experiments. The remaining two subjects prematurely terminated the study of their own accord after the high dose psilocybin session. Both subjects reported having had experiences of strong anxiety, fear of loss of ego control, emerging negative memories, and thoughts during acute drug effects and were therefore not willing to participate in further psilocybin sessions. All five adverse drug reactions leading to a premature termination of the study were confined to the acute phase of drug effects and were completely resolved by the end of the experimental day.

3.4.2 *Acute Psychological Effects*

3.4.2.1 *Dimensions of ASCs*

Psilocybin significantly increased scores of all 5D-ASC scales [main effects of drug in order of significance: VRS: $F(5, 219) = 100.47$, $p < 0.001$; G-ASC: $F(5, 219) = 89.34$, $p < 0.001$; OBN: $F(5, 219) = 59.93$, $p < 0.001$; VIR: $F(5, 219) = 47.08$, $p < 0.001$; AUA: $F(5, 219) = 23.01$, $p < 0.001$; DED: $F(5, 219) = 21.53$, $p < 0.001$]. Dose-dependent effects of psilocybin on the sub-scale level are illustrated in Figure 2. Means and standard deviations of each 5D-ASC in each dose condition is available in Supplementary Table S1. One-tailed Dunnett's contrasts detected significant differences between placebo and all drug dose conditions in all 5D-ASC scales, except for the 45 µg/kg dose condition, which was not significantly different from placebo in any of the 5D-ASC scales.

Graphical representations of drug effects by locally weighted scatter plot smoothing curves indicated dose-response-relationships that were reasonably well approximated by linear functions in all scales of the 5D-ASC. The notion of linear dose-response relations was also supported by the results of orthogonal polynomial contrasts, which showed highly significant effects for the linear trends in all 5D-ASC scales. Furthermore, when dose was treated as a continuous variable in the linear mixed effects models, regression slopes for doses were significantly different from zero in all 5D-ASC scales. The regression slope for dose was highest with the dependent variable VRS (estimate ± SE; 0.154 ± 0.010) followed by OBN (0.117 ± 0.008), VIR (0.100 ± 0.014), DED (0.079 ± 0.010), and AUA (0.047 ± 0.010). Thus, for an increase of 100 µg/kg of psilocybin, the regression model predicted an average increase of 15% in the VRS, 12% in the OBN, and 8% in the DED scale.

Cumulative distributions of the three aetiology independent dimensions of ASCs for each of four doses of psilocybin and placebo are displayed in Figure 3. As can be seen from the plot, the cumulative distributions for all psilocybin doses were relatively smooth over the whole ranges of the response variables indicating widely varying individual responses. The proportions of subjects experiencing strong drug effects (scores > 70% of the maximum possible score) were clearly dose dependent. In the highest dose condition, 22% of subjects reached or exceeded the cut-off value for strong OBN, whereas only 5.7%, 7.8%, and 0% of subjects experienced such a strong effect in the 250-260, 215, and 115-125 µg/kg psilocybin dose conditions, respectively. Experiences of pronounced DED occurred in relatively few subjects and were only observed in the two highest dose conditions. Specifically, 7.3% and 5.7% of subjects reached or exceeded the

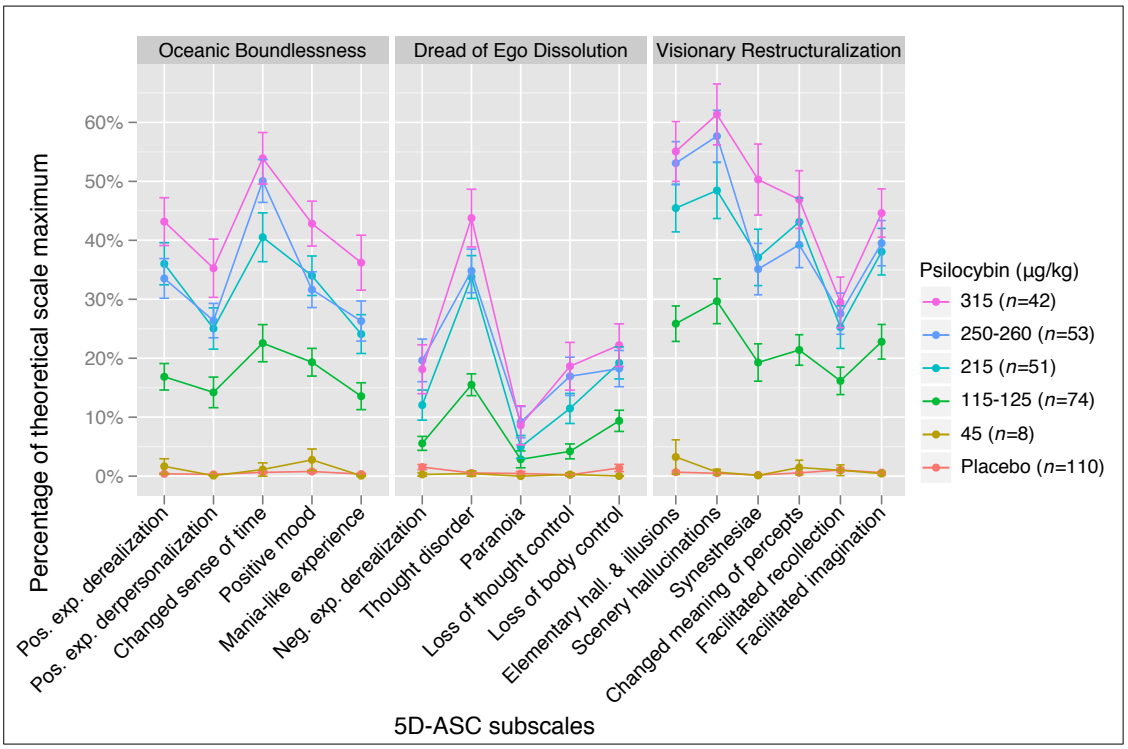

Figure 2: Dose-dependent percentage scores of item clusters from the Five Dimensions of Altered States of Consciousness (5D-ASC) rating scale. Error bars represent standard errors. Ratings were obtained during peak drug effects (60-270 min after drug administration). Double click here to extract the data underlying this plot.

cut-off-value for DED in the 315 and 250-260 µg/kg condition, respectively. Finally, the number of subjects exhibiting high to very high VRS scores amounted to 19.6%, 7.5%, 7.8%, and 1.3% in the 315, 250-260, 215, and 115-125 µg/kg dose conditions, respectively.

3.4.2.2 *Eigenschaftswörterliste 60-S (EWL-60-S)*

Psilocybin induced changes in affective mood states and conditions over four different time periods are summarized in Figure 4. Mixed effects models fitted for each of the 15 EWL-60-S subscales revealed significant Drug × Time interactions for emotional excitation [$F(3, 284 = 12.54$, $p < 0.001$], dreaminess [$F(3, 284) = 11.62$, $p < 0.001$], heightened mood [$F\ (3, 284) = 9.23$, $p < 0.001$], dazed state [$F(3, 284) = 6.59$, $p = 0.005$], sensitivity [$F(3, 284) = 5.64$, $p = 0.015$], concentration [$F(3, 284) = 4.98$, $p = 0.032$], and tiredness [$F\ (3, 284) = 4.74$, $p = 0.041$]. Interaction plots indicated that all significant interaction effects, except for tiredness, were produced by stronger drug effects on early measuring time points relative to later time points and hence represent acute drug effects. The interaction effect for tiredness was produced by stronger drug effects on later measuring time points relative to earlier time points and hence represents a psilocybin after effect.

Significant main effects of drug in the absence of significant Drug × Time interactions were detected for introversion [$F(3, 284) = 60.96$, $p < 0.001$], inactivation [$F(3, 284) = 50.82$, $p < 0.001$], efficiency-activation [$F(3, 284) = 38.92$, $p < 0.001$], extroversion [$F(3, 284) = 12.40$, $p = 0.014$], and apprehension-anxiety [$F(3, 288) = 12.36$, $p < 0.001$]. Since

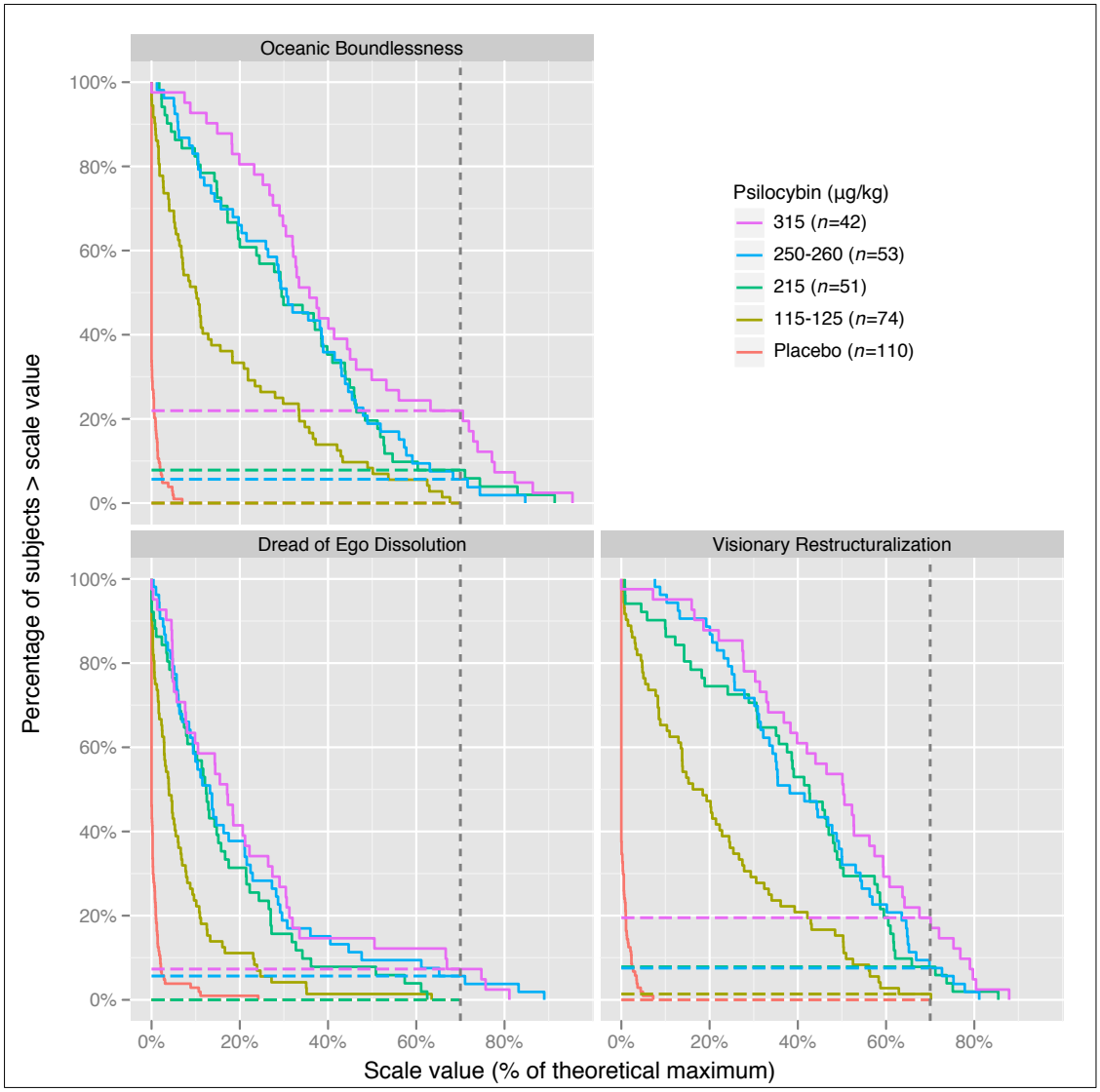

Figure 3: Cumulative distributions of the three aetiology-independent dimensions of the Five Dimensions of Altered States of Consciousness (5D-ASC) rating scale for each of four psilocybin doses and placebo. Dashed reference lines mark strong subjective drug effects (70% of theoretical scale maxima). Double click here to extract the data underlying this plot.

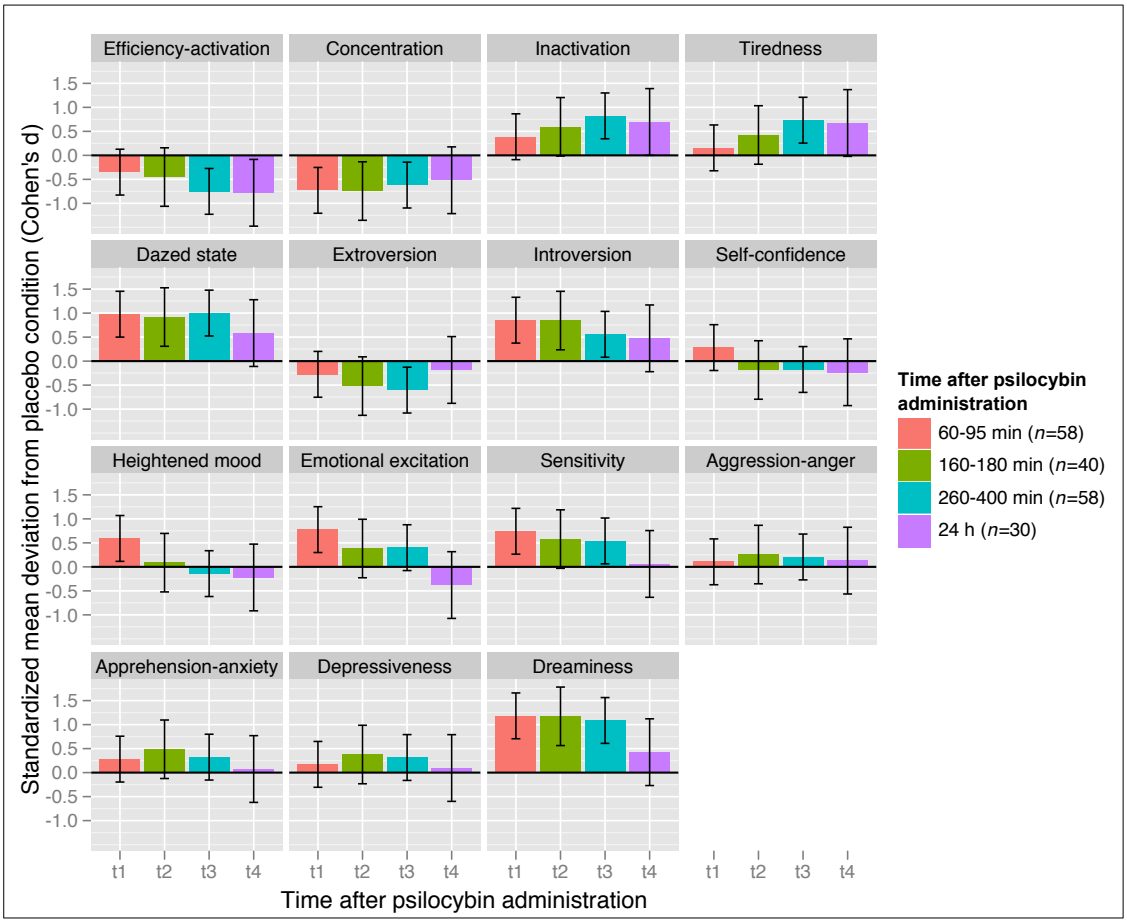

Figure 4: Time-dependent effects of medium-dose psilocybin (215-260 µg/kg) on mood states and condition measured by the Eigenschaftswörterliste 60-S (EWL-60-S). Paired differences between placebo and psilocybin conditions for each time and variable combination were divided by their standard deviations in order to express psilocybin effects in units of Cohen's *d* effect size. By convention, Cohen's *d* of 0.2, 0.5, and 0.8 are termed small, medium, and large effect sizes, respectively. Error bars denote Bonferroni-corrected 95% confidence intervals of mean differences. Thus, mean differences between placebo and psilocybin are significant, where error bars do not include zero. Double click here to extract the data underlying this plot.

these drug-effects were independent of measurement time, they represent longer-lasting psilocybin-effects.

3.4.3 *Short-Term Side Effects*

Global scores of the List of Complaints measured 24 h after drug intake were dependent on drug condition [$F(5, 159) = 9.64$, $p < 0.001$]. All pair-wise comparisons between placebo and active psilocybin conditions were statistically significant except for the very low dose psilocybin condition. No comparison between any of two active psilocybin dose conditions was statistically significant. Single complaints registered 24 h after psilocybin administrations are summarized in Table 2. Item-level comparisons between three different psilocybin doses and placebo by Cochran Q tests revealed significant

Table 2: List of Complaints, 24 h After Drug Intake

	Dose effect relation (n = 40)						Medium dose comparison (n = 72)			
	Placebo	115 µg/kg	215 µg/kg	315 µg/kg	p-value[a]	Signif.	Placebo	215-250 µg/kg	p-value[b]	Signif
Fatigue	12.5% (5)	40.0% (16)	35.0% (14)	60.0% (24)	< 0.001	***	40.3% (29)	19.4% (14)	0.009	**
Exhaustion	7.5% (3)	22.5% (9)	22.5% (9)	22.5% (9)	0.090		27.8% (20)	9.7% (7)	0.009	**
Headaches, head pressure or face pain	2.5% (1)	12.5% (5)	22.5% (9)	37.5% (15)	< 0.001	***	19.4% (14)	8.3% (6)	0.099	
Lack of energy	0.0% (0)	15.0% (6)	7.5% (3)	22.5% (9)	0.002	**	16.7% (12)	4.2% (3)	0.027	*
Excessive sleep requirement	2.5% (1)	10.0% (4)	10.0% (4)	15.0% (6)	0.177		12.5% (9)	6.9% (5)	0.386	
Difficulty concentrating	5.0% (2)	7.5% (3)	7.5% (3)	17.5% (7)	0.015	*	13.9% (10)	4.2% (3)	0.046	*
Gone feeling	2.5% (1)	10.0% (4)	5.0% (2)	22.5% (9)	0.005	**	12.5% (9)	2.8% (2)	0.023	*
Fast exhaustibility	2.5% (1)	12.5% (5)	10.0% (4)	17.5% (7)	0.064		8.3% (6)	4.2% (3)	0.371	
Brooding	5.0% (2)	5.0% (2)	0.0% (0)	12.5% (5)	0.106		12.5% (9)	4.2% (3)	0.114	
Lack of appetite	0.0% (0)	7.5% (3)	5.0% (2)	17.5% (7)	0.015	*	9.7% (7)	1.4% (1)	0.077	
Neck or shoulder pain	7.5% (3)	7.5% (3)	2.5% (1)	5.0% (2)	0.629		8.3% (6)	4.2% (3)	0.450	
Irritability	5.0% (2)	10.0% (4)	5.0% (2)	7.5% (3)	0.768		5.6% (4)	2.8% (2)	0.683	
Sexually stimulating phantasies	5.0% (2)	2.5% (1)	5.0% (2)	5.0% (2)	0.801		5.6% (4)	6.9% (5)	1.000	
Strong thirst	2.5% (1)	5.0% (2)	5.0% (2)	0.0% (0)	0.468		9.7% (7)	1.4% (1)	0.077	
Heavy or tired legs	2.5% (1)	0.0% (0)	2.5% (1)	12.5% (5)	0.008	**	4.2% (3)	2.8% (2)	1.000	
Sleeplessness	2.5% (1)	0.0% (0)	7.5% (3)	5.0% (2)	0.290		6.9% (5)	1.4% (1)	0.221	
Bloated feeling	5.0% (2)	0.0% (0)	5.0% (2)	2.5% (1)	0.300		5.6% (4)	2.8% (2)	0.617	
Backache	2.5% (1)	5.0% (2)	2.5% (1)	2.5% (1)	0.896		5.6% (4)	2.8% (2)	0.683	
Worries about prof. or private affairs	0.0% (0)	5.0% (2)	2.5% (1)	5.0% (2)	0.532		4.2% (3)	4.2% (3)	1.000	
Dark thoughts	5.0% (2)	5.0% (2)	2.5% (1)	2.5% (1)	0.801		2.8% (2)	4.2% (3)	1.000	
Inner tension	2.5% (1)	7.5% (3)	0.0% (0)	7.5% (3)	0.234		2.8% (2)	1.4% (1)	1.000	
Abdominal pain or stomach ache	2.5% (1)	2.5% (1)	2.5% (1)	7.5% (3)	0.392		2.8% (2)	1.4% (1)	1.000	
Intolerances to certain smells	0.0% (0)	2.5% (1)	2.5% (1)	5.0% (2)	0.494		5.6% (4)	1.4% (1)	0.371	
Nausea	0.0% (0)	7.5% (3)	5.0% (2)	2.5% (1)	0.232		4.2% (3)	0.0% (0)	0.248	
Uneasiness	2.5% (1)	2.5% (1)	2.5% (1)	5.0% (2)	0.875		2.8% (2)	1.4% (1)	1.000	
Tendency of crying	2.5% (1)	0.0% (0)	2.5% (1)	0.0% (0)	0.572		5.6% (4)	1.4% (1)	0.371	
Joint aches	2.5% (1)	2.5% (1)	2.5% (1)	0.0% (0)	0.733		4.2% (3)	1.4% (1)	0.617	
Cold feet	2.5% (1)	0.0% (0)	2.5% (1)	2.5% (1)	0.801		4.2% (3)	1.4% (1)	0.617	
Freezing	0.0% (0)	2.5% (1)	5.0% (2)	0.0% (0)	0.300		4.2% (3)	1.4% (1)	0.617	
Ravenous appetite	5.0% (2)	2.5% (1)	2.5% (1)	0.0% (0)	0.494		1.4% (1)	2.8% (2)	1.000	
Throat pain or irritated throat	5.0% (2)	0.0% (0)	0.0% (0)	7.5% (3)	0.101		0.0% (0)	2.8% (2)	0.480	
Easy rubescence	2.5% (1)	7.5% (3)	0.0% (0)	0.0% (0)	0.066		0.0% (0)	2.8% (2)	0.480	
Lump in throat or throat tightness	2.5% (1)	2.5% (1)	2.5% (1)	0.0% (0)	0.733		1.4% (1)	1.4% (1)	1.000	
Diarrhoea	2.5% (1)	0.0% (0)	0.0% (0)	7.5% (3)	0.112		0.0% (0)	1.4% (1)	1.000	
Restless legs	5.0% (2)	2.5% (1)	0.0% (0)	0.0% (0)	0.300		0.0% (0)	2.8% (2)	0.480	
Cold intolerance	0.0% (0)	2.5% (1)	2.5% (1)	0.0% (0)	0.392		4.2% (3)	0.0% (0)	0.248	
Vertigo	2.5% (1)	0.0% (0)	0.0% (0)	2.5% (1)	0.392		1.4% (1)	2.8% (2)	1.000	
Forgetfulness	0.0% (0)	0.0% (0)	2.5% (1)	5.0% (2)	0.194		2.8% (2)	0.0% (0)	0.480	
Difficulty swallowing	2.5% (1)	0.0% (0)	0.0% (0)	5.0% (2)	0.300		0.0% (0)	1.4% (1)	1.000	
Frequent urges to urinate	2.5% (1)	0.0% (0)	2.5% (1)	0.0% (0)	0.572		1.4% (1)	1.4% (1)	1.000	
Strong perspiration	2.5% (1)	2.5% (1)	0.0% (0)	0.0% (0)	0.572		1.4% (1)	1.4% (1)	1.000	

Numbers in parenthesis indicate absolute frequencies. Complaints are ordered by row sums of absolute frequencies. Complaints with an absolute frequency < 4 over all drug conditions are not shown in the table. [a]Cochran Q-tests. [b]McNemar tests.

differences for the items fatigue, headaches, lack of energy, difficulty concentrating, "gone feeling", lack of appetite, and heavy or tired legs.

Item-level comparisons of the List of Complaints responses in the largest possible sample (i.e., medium dose psilocybin vs. placebo) by McNemar tests indicated increased fatigue, exhaustion, lack of energy, difficulty concentrating, and "gone feeling" after psilocybin administration. However, if a Holm-correction for multiple comparisons is applied, only the items fatigue, headaches, and lack of energy in the dose-effect comparison remain statistically significant. Serious complications, such as fear of death, shortness of breath, feelings of suffocation, vomiting, and fainting, which were also covered by the questionnaire, were not reported by any of the subjects.

3.4.4 *Long-Term Follow-Up*

Long-term follow-up questionnaires were completed by 90 of 110 subjects (82%). 20 subjects were either unavailable due to address change or unresponsive. Chi-square and Welch's t tests indicated that those subjects who completed the follow-up questionnaire were not statistically different from subjects who did not with respect to age, gender, education, 5D-ASC, and list of complaints scores. Subjects completed follow-up assessments between 8 and 16 months after their last experimental day (M $\pm$ SD: 330 $\pm$ 90 days).

3.4.4.1 *Retrospective ratings of acute drug effects and changes in values and attitudes*

Retrospective ratings of acute drug effects are summarized in Table 3, whereas changes in values and attitudes are shown in Table 4.

3.4.4.2 *Changes in drug consumption habits*

Changes in consumption habits of the most often used psychotropic substances are summarized in Table 5. Most subjects reported unchanged consumption habits for all drugs. Those subjects who did report changes more often described decreased consumption. Even for psilocybin itself, more subjects reported to have consumed it less often (5.6%) than more often (3.3%). Of the three subjects who described increased psilocybin consumption, two subjects reported to consume it twice a year and one three times per year. Except for alcohol, nicotine, and cannabis, no drug was used more often than once per month on average. From the twenty-nine subjects who reported changes, seven subjects (24%) considered the change as a direct consequence of their hallucinogen experience. One of these seven reported decreased substance consumption, another reported increased, and five reported both increased and decreased.

3.4.4.3 *Spontaneous alterations of consciousness and flashbacks*

Nine subjects (10%) reported spontaneously occurring ASCs before and eight (9%) after the experiments. Three of these subjects reported experiencing spontaneous ASCs both

Table 3: Acute Drug Effects

Adjective	Subjects $n = 90$
pleasant	
very	49% (43)
medium	43% (38)
no	8% (7)
enriching	
very	61% (53)
medium	29% (25)
no	10% (9)
frightening	
very	5% (4)
medium	28% (24)
no	68% (59)
unpleasant	
very	10% (9)
medium	24% (21)
no	66% (57)
influential	
very	22% (19)
medium	45% (39)
no	33% (29)
nothing special	
very	4% (3)
medium	11% (9)
no	85% (70)

Numbers after percents are frequencies.

before and after the experiments. Spontaneous ASCs before the experiments included out-of-body-experiences during meditation and sleep, trance-like states while deeply concentrating, euphoric experiences in nature, perceptual alterations in very dark or bright environments, lucid dreams, hearing voices under high fever, and hypnagogic hallucinations. All these alterations lasted a few seconds to no more than one hour, occurred a few times in a lifetime to no more than once per month, and were limited to specific triggers. They were not experienced as threatening and they did not interfere with subjects' everyday lives. Thus, they cannot be interpreted as psychopathological symptoms.

Spontaneous ASCs after the experiments could not be distinguished from those before with respect to frequency, duration, and intensity. Except for one subject who reported irritability, depressive feelings, anxiety/panic, and dizziness/nausea and who will be discussed in more detail below, all spontaneous ASCs that occurred after the experiments were described as non-threatening and without impairment in social, occupational, or

Table 4: Changes in Attitudes

Change	Subjects $n = 90$
Changes in world view	
positive	18% (16)
unchanged	81% (72)
negative	1% (1)
Changes in values	
positive	18% (16)
neither	77% (68)
negative	5% (4)
Changes in awareness of personal problems	
positive	29% (25)
unchanged	66% (57)
negative	5% (4)
Change in the relationship to one's body	
positive	30% (26)
unchanged	67% (59)
negative	3% (3)
Change in relationhips to other people	
positive	25% (22)
unchanged	68% (59)
negative	7% (6)
Change in professional relationships	
positive	6% (5)
unchanged	89% (75)
negative	5% (4)
Change in the relationship to the environment	
positive	38% (33)
unchanged	58% (51)
negative	5% (4)
Change in aesthetic experiencing	
positive	37% (32)
unchanged	62% (53)
negative	1% (1)
Change in the attitude to ASCs	
positive	56% (49)
unchanged	41% (36)
negative	3% (3)

Numbers after percents are frequencies.

Table 5: Change in Drug Consumption

Drug	less often	more often
Alcohol	6.7% (6)	3.3% (3)
Nicotine	4.4% (4)	2.2% (2)
Cannabis	8.9% (8)	3.3% (3)
MDMA (Ecstasy)	4.4% (4)	3.3% (3)
Psilocybin	5.6% (5)	3.3% (3)
Cocaine	4.4% (4)	1.1% (1)
Amphetamine	2.2% (2)	0% (0)

Numbers after percents are frequencies.

other important areas of functioning. They were controllable and limited to specific triggers, such as listening to music, meditating, falling asleep (hypnagogic states), deeply concentrating, or being in a calm and sensory-deprived environment. In all subjects, spontaneous ASCs after the experiments were described as lasting a few seconds to no more than half an hour and occurring infrequently.

Although three subjects reported minor visual alterations, they mostly appeared after the triggers mentioned above. Moreover, the descriptions of these alterations were vague and not suggestive of the typical symptoms of Hallucinogen Persisting Perception Disorder (HPPD) mentioned in DSM-IV (i.e., geometric hallucinations, false perceptions of movement in the peripheral visual fields, flashes of color, intensified colors, trails of images of moving objects, after-images of moving objects, macropsia, and micropsia). Only one subject explicitly mentioned the occurrence of visual illusions beyond the acute effects of the drug, but they appeared only 5-7 times, lasted no more than a few seconds at a time, and did not occur after the third day after the drug session. When asked whether they interpreted their spontaneous ASC as a flashback-like re-experiencing of drug effects, five subjects (5.5% of those who completed the follow-up questionnaire and 56% of those who reported spontaneous ASCs after the experiments) answered affirmatively. However, since we have also obtained detailed descriptions of these events, including frequency, duration, intensity, and accompanying symptoms and emotions, there was a clear indication that these five subjects had used the term flashback in a very broad sense, denoting vague states of intense remembering of drug effects rather than the criteria described for HPPD in DSM-IV.

3.4.4.4 *Negative changes in psychological well-being and/or mental functions*

Eleven subjects (12%) reported in the follow-up questionnaire that they had experienced negative changes in psychological well-being and/or mental functions after the psilocybin experiment. However, four of these eleven reported that the changes were unrelated to the psilocybin sessions and so they will not be discussed here. Among the remaining seven subjects (8%), only one reported that his symptoms were alarming and severe enough for him to contact us and to seek professional help. This was the same subject who mentioned irritability, anxiety, and depressive feelings in the items regarding spontaneous ASCs described above. This subject was a 23 year-old medical

student who presented himself as psychologically stable and with an unremarkable medical history at screening. There was also no indication of above average emotional lability according to the FPI questionnaire. During the acute effects of the high psilocybin dose (315 µg/kg), the subject experienced a strong sense of unity, but also intense feelings of loneliness and fear of losing control of his thoughts and body. 150 min after drug intake, he reached an DED score of 76% of the theoretical scale maximum, which is the fourth highest score that we measured in the 227 psilocybin sessions of the present analysis. After providing strong personal support and reassurance by the study manager, the subject had calmed down considerably by the second 5D-ASC measuring time point (300 min after drug intake), reaching only 4% of the maximum possible score in the DED scale. Nevertheless, the subject was monitored carefully and not released until 8 h after drug intake when all acute psilocybin effects had fully worn off. However, since the subject felt uncomfortable over the next couple of weeks, several appointments were made with the principal investigator. In these meetings, the subject reported emotional instability, anxiety, and depressive feelings, which he attributed to suppressed memories, thoughts, and feelings that he had been confronted with during the psilocybin session. Since the subject was strongly motivated to work through these issues psychologically, he was referred to an experienced psychotherapist. After a few sessions of psychotherapy, the subject had completely stabilized and has not relapsed with the symptoms subsequently.

Among the remaining six subjects who also reported negative changes in well-being and/or mental functions, the following symptoms were described (with number of subjects in parenthesis): Concentration problems (2), mood swings (2), reactivation of old problems (1), memory problems (1), and being pensive and introverted (1). In all these subjects, symptoms were described to be of low intensity and frequency, non-interfering with everyday life, and only occurring in the first few weeks subsequent to the experiments.

3.5 DISCUSSION

The present work analyzed acute, short-, and long-term effects of psilocybin in healthy human subjects by pooling raw data from a large body of well-controlled experimental studies.

3.5.1 *Acute Psychological effects*

Consistent with other recent, smaller-scaled studies (Griffiths, Richards, McCann, & Jesse, 2006; Moreno et al., 2006), the present analysis has shown that psilocybin dose-dependently induces an ASC, which is characterized by marked alterations in all mental functions, including perception, mood, volition, cognition, and self-experience. The most prominent features of the psilocybin-induced ASCs were alterations in visual perception (VRS), followed by positively (OBN) and negatively (DED) experienced alterations of self-awareness and loosening of ego-boundaries. The changes in visual perception ranged from increased visual imagery with closed eyes, optical illusions, elementary hallucinations, and synesthesiae to picture-like scenery hallucinations. However, the experienced hallucinations were almost always recognized as unreal and therefore are

more accurately described as pseudo- or nonpsychotic hallucinations. Drug effects accounting for the substantial increase in the OBN scale, ranged from pleasurable experiences of depersonalization, derealization, and a changed sense of time to phenomena reminiscent of mystical-type experiences. The modest increase in the DED scale was primarily due to unpleasant disturbances of cognitive functions and somatesthesia, and much less so to suspiciousness or paranoid ideation. Reality testing usually remained intact, and most subjects sustained critical distance ("it is as if") to their own subjective experience. The Auditory Alterations (AUA) scale was only moderately affected by psilocybin, since true auditory hallucinations, such as hearing voices, rarely occurred, and auditory alterations mostly concerned occasional intensification of music and sounds or misperceptions of real auditory stimuli. Psilocybin also dose-dependently increased the Vigilance Reduction (VIR) scale. This observation may come as a surprise, since lack of sedation and clouding of consciousness is usually considered as one of the most prominent characteristics of classical hallucinogens. In fact, according to the classification scheme of Leuner (1981), hallucinogens of the first order (e.g., LSD, psilocybin, DMT, and mescaline) are differentiated from hallucinogens of the second order (e.g., ketamine, N_2O, and scopolamine) by this very feature. However, it should be noted that the effect of psilocybin on reduction of vigilance was relatively small and reflects the psilocybin-induced state of dreaminess, contemplativeness, and reduction of attentiveness, rather than true sedation or clouding of consciousness.

Cumulative distributions of the 5D-ASC major scales revealed widely varying individual responses. For instance, whereas one subject experienced strong effects on the low dose condition (63% of the possible maximum of the global scale), two subjects noticed almost no effects on the highest dose condition (below 5% of the possible maximum). The high inter-subject and moderate inter-study variability of the pooled analysis supports the view that psilocybin effects are poorly predicted by drug dose alone and that other pharmacological variables, such as plasma levels of the active metabolite psilocin, as well as non-pharmacological variables - notably expectations, personality structure, interpersonal support, and environment - likely play a very important role (Rinkel, DiMascio, Robey, & Atwell, 1961; Metzner, Litwin, & Weil, 1965; Dittrich, 1994; Johnson, Richards, & Griffiths, 2008). The significance of such potential non-pharmacological predictors will be analyzed and presented in a separate publication.

Dose-response relationships for all major scales of the 5D-ASC were approximately linear. Since no ceiling effect has been observed within the administered dose range (45-315 µg/kg body weight), it is conceivable that psilocybin doses exceeding 315 µg/kg would have produced even stronger subjective drug effects. This view is supported by a recent study of Griffiths, Richards, McCann, and Jesse (2006), in which subjective drug effects were measured by the APZ questionnaire in 36 healthy volunteers who had received 429 µg/kg psilocybin. By estimating 5D-ASC from APZ scores through linear equations (Bodmer, 1999), we have found higher OBN and VRS scores in the study of Griffiths than in the highest dose condition of our studies. Furthermore, in the study of Griffiths, 61% of subjects fulfilled Pahnke's criteria for having a 'complete' mystical experience (Pahnke, 1969), which seems to be a considerably larger proportion than we have observed in the highest dose condition. Although we have used a different methodology and therefore cannot directly compare results, we have found that only 22% of subjects in the high-dose condition exceeded the cut-off value of the OBN scale

suggestive of deep mystical or transcendent experiences. However, it should it be noted that, in addition to the higher drug doses in the study of Griffiths, several other factors might have contributed to these differences. First, the investigation of transformative peak experiences has not been the primary goal of our research program. Hence, our studies have not been designed in a way that the occurrence of such profound experiences is most likely. Whereas volunteers in the study of Griffiths were instructed to focus explicitly on the phenomenology of the drug experience and were left essentially undisturbed during their whole psilocybin session, subjects of our studies were engaged in performing tasks for a considerable amount of time. Second, none of the subjects in Griffith's study had previous experience with a hallucinogenic drug, whereas in our studies about 40% of subjects had previous experience with a classical hallucinogen (LSD, psilocybin, DMT, or mescaline) and almost 90% had smoked cannabis at least once in a lifetime. Third, the subjects in Griffith's study were middle-aged (46 years on average) and spiritually active, whereas our subjects were predominantly students, considerably younger (27 years on average), and not selected for being spiritual.

Despite extremely careful preparation, selection, and interpersonal support of subjects, there also seems to have occurred more acute adverse reactions in the study of Griffiths than in our studies. Griffiths reports that 31% of subjects experienced significant fear and 17% had transient ideas of reference/paranoia, whereas in our studies only 7% of subjects in the highest dose condition fulfilled the criteria for strong dread of ego dissolution, which is suggestive of acute psychotic reactions. Although these adverse reactions were confined to the acute phase and were readily managed by providing interpersonal support without psychopharmacological intervention in all cases of both research groups, they provide a cautionary note on high-dose psilocybin studies. Clearly, careful selection of subjects and environment and thorough preparation and monitoring of subjects is extremely critical in high-dose psilocybin sessions.

The results of our analysis of the time-course of subjective effects measured by the EWL-60-S indicate that the effects follow differential time courses, which are not necessarily paralleled by psilocybin plasma levels (Hasler, Bourquin, Brenneisen, Bär, & Vollenweider, 1997). Whereas the effects of psilocybin on emotional excitation, sensitivity, heightened mood, and concentration reached their maximum in an early phase (60-180 min after drug intake), the effects on dreaminess, dazed state, inactivation, and introversion were more pronounced in a later phase (260-400 min). The results are consistent with a study of Heimann (1961), in which expressive phenomena, such as changes in facial expression, voice, and posture, were analyzed by use of video recording in 12 healthy volunteers who had received 0.06-0.19 μg/kg psilocybin on two experimental days. Heimann observed that subjects were more active, emotional, vivid, extroverted, and cognitively impaired in the early phase of the psilocybin session relative to the later phase and that derealization and depersonalization phenomena began to dominate over visual alterations about 90-120 min after drug intake. During this later phase, subjects also increasingly turned inwards and appeared to be in a state of absent-mindedness with markedly reduced facial expression.

3.5.2 *Short-Term Side Effects*

Our pooled analysis further revealed that the administration of psilocybin caused only few subacute side-effects, as measured by the LC questionnaire 24 h after drug intake. Furthermore, those complaints that were reported significantly more often after psilocybin than after placebo concerned relatively mild conditions. Except for headaches, all significantly affected items described symptoms of tiredness and exhaustion. Serious complications were not reported in any of the subjects. The reported after-effects of the LC questionnaire closely match the changes seen in the EWL-60-S questionnaires 24 h after drug intake, namely, a moderate increase of tiredness and reduction of activation. The results of our analysis therefore suggest that psilocybin is usually well tolerated and that normal functioning is almost completely restored within 24 h after drug administration. These findings are in line with an investigation of Hollister (1961), which measured after effects of the administration of 36-205 µg/kg psilocybin to 17 subjects in 27 separate trials and also found occasional headaches and fatigue as being the most frequent complaints. Furthermore, our results are consistent with extensive tests in animals and humans, which found that psilocybin may be considered to be physiologically well tolerated (for a review, see Passie, Seifert, et al., 2002).

3.5.3 *Long-Term Follow-Up*

3.5.3.1 *Retrospective ratings of acute drug effects*

At the time of the long-term follow-up 8-16 months after the last experimental session, the majority of subjects were still positively impressed by the psilocybin experience. When subjects were asked to rate the acute psilocybin effects by six descriptive items, "enriching" was considered as most applicable. Over 60% of subjects rated the experience as very enriching and over 90% as enriching to at least a medium degree. Interestingly, several of our subjects rated the psilocybin experience as very enriching even though they had experienced significant distress during the acute phase. For instance, among the sixteen subjects who had an DED score of more than 50% of the maximum possible score, four subjects (25%) did not respond to the follow-up questionnaire, nine subjects (75% from those who responded) rated the experience as very enriching, two (17%) as medium enriching, and only one subject (8%) as not enriching. The positive long-term resolution of acute distressful experiences by the majority the subjects might be partially explained by the strong support provided by our monitors, with whom subjects were able to talk freely about disturbing thoughts, feelings, and memories that had arisen during the session. A significant number of our subjects reported not only enriching but also influential drug experiences. Our results are therefore in support of a recent follow-up study by Griffiths, Richards, Johnson, et al. (2008), in which psilocybin was found to facilitate experiences having enduring personal meaning and spiritual significance.

3.5.3.2 *Changes in values and attitudes*

Anecdotal reports and preliminary evidence from small-scaled experimental studies suggest that hallucinogenic drugs, when used under carefully controlled and supportive

conditions, sometimes can lead to sustained positive changes in personality, attitudes, and values, particularly in those subjects who have experienced profound personal insights and transcendent or mystical-type experiences. Among the most often reported subjective changes in attitude and personality are more self-understanding, more tolerance of others, less egocentricity, a less materialistic and aggressive orientation, and more appreciation of music, art, and nature (McGlothlin & Arnold, 1971). Subjective changes in attitudes reported in the long-term follow-up questionnaire of the present study are consistent with results of earlier follow-up studies (McGlothlin, Cohen, & McGlothlin, 1967; Doblin, 1991; Griffiths, Richards, Johnson, et al., 2008). We have found the highest percentages of subjects reporting positive changes in items measuring the attitude to ASCs (56% of subjects), relations to the environment/nature (38%), and aesthetic experiencing (37%). The observation that aesthetic experiencing (e.g., enhanced appreciation of art and music) was amongst the most often reported positive changes is particularly interesting in light of an earlier placebo-controlled study by McGlothlin, Cohen, and McGlothlin (1967), which measured long-lasting effects of three high-dose LSD sessions in 24 healthy volunteers. In this study, greater appreciation of music (67% of subjects) and art (46%) were the most frequently reported subjective changes six months after the LSD experiments. Furthermore, these subjective evaluations were supported by certain behavioral changes, such as increase in number of records bought, time spent in museums, and number of musical events attended.

Given the positive changes in attitudes and values reported by a relatively large proportion of subjects, it would be tempting to conclude that hallucinogenic drugs hold a large and presently unused potential for increasing life-satisfaction and personal growth and for assisting psychotherapy. However, our results should be considered as exploratory in nature because possible changes have not been measured by validated questionnaires and no attempt has been made to correlate subjective changes with behavioral measures or information provided by close relatives and friends. Therefore, we cannot rule out the possibility that the reported changes are biased towards preconceived opinions and expectations of subjects. Caution is especially warranted since it has been demonstrated that positive changes in personality, attitudes, and values are often attributed to the hallucinogen experience in subjects who have shown a previous interest in hallucinogenic drugs, but not in subjects whose hallucinogen intake was initiated by their psychotherapist (McGlothlin & Arnold, 1971).

3.5.3.3 *Changes in drug consumption habits*

Changes in drug use patterns that were reported in the follow-up questionnaires of the present study were generally benign and well within expected ranges. Our results therefore indicate that a carefully-monitored administration of 1-4 doses of psilocybin to healthy volunteers within an experimental setting does not increase the risk for subsequent abuse of psilocybin or other illicit drugs. Our results are consistent with the widely accepted view that classical hallucinogens have a very low abuse potential. Classical hallucinogens are not typically considered as drugs of addiction because they neither produce compulsive drug-seeking behavior nor physical withdrawal symptoms (O'Brian, 2005). They also cannot be considered as reinforcing substances because they fail to engender reliable self-administration behavior in laboratory animals (Deneau, Yanagita, & Seevers, 1969; Fantegrossi, Murnane, & Reissig, 2008). The view that

classical hallucinogens lack addictive qualities is further supported by epidemiological evidence. Recent general population survey data on lifetime prevalence of use of hallucinogenic mushrooms in 12 EU Member States indicate that among young people aged 15 to 24 years old, lifetime use of hallucinogenic mushrooms ranges from less than 1% to 8% (Hillebrand, Olszewski, & Sedefov, 2006). Although psilocybin has, after cannabis, one of the highest lifetime prevalence rates of all illicit drugs, the proportion of recent (last 12 months) or current (last month) users is much lower for the use of psilocybin than it is for cannabis and ecstasy. This observation suggests that the use of hallucinogenic mushrooms, like LSD, tends to be occasional, or discontinued after some time. Regular use of classical hallucinogens is unlikely because tolerance to the effects rapidly develops after three to four daily doses (O'Brian, 2005). Furthermore, the intake of classical hallucinogens, especially in higher doses, is unattractive to many recreational drug users because it does not consistently produce any of the pleasurable effects of addictive drugs, such as escape, euphoria, anxiety relief, increase of self-esteem etc. Although our experiments have shown that psilocybin can evoke highly-valued and in some cases even mystical-type experiences, subjects occasionally are also confronted with frightening and unpleasant thoughts, memories, and emotions. Moreover, as we have measured by the LC and EWL-60-S questionnaires, most subjects describe the psilocybin effects as tiring. This effect is further reflected by the observation that after 3-5 h the 'coming-down' from the psilocybin-effects, even if it has been a pleasant experience, is usually welcomed, and that most subjects are glad to regain their normal state of consciousness. Subjects often reported they were 'saturated' by new impressions and expressed the need to psychologically integrate their experience before they would consider repeating it.

Our findings are in line with a 10-year follow-up study by McGlothlin and Arnold (1971), which examined 247 subjects who had received LSD in either an experimental or therapeutic setting. As in our study, most subjects reported to have discontinued or reduced their frequency of hallucinogenic drug use. The most often reported reasons for discontinuation were concerns about possible harm or illegality followed by a loss of interest. McGlothlin therefore speculated that in many subjects, hallucinogenic drugs lose their appeal over time simply because the uniqueness of the new modes of perception and thought that occur with them being often the primary incentive to take them in the first place is lost after repeated ingestion.

3.5.3.4 *Spontaneous alterations of consciousness and flashbacks*

Among the most often reported long-term sequelae of hallucinogenic drug use is a sudden and unexpected reoccurrence of all or certain aspects of the hallucinogenic effects, long after the drug should have worn off. The phenomenon has been first described by (Sandison, 1954) and is often referred to as 'flashback'. However, since its first description in the scientific literature (Horowitz, 1969), the term 'flashback' has been defined in so many ways that much confusion exists about its characteristics, prevalence, and aetiology (cf. Halpern & Pope, 2003). The present study avoided methodological inconsistencies of earlier follow-up studies and contributes to a better understanding of flashback phenomena after psilocybin administration by using operational criteria consistent with those of DSM-IV ('Hallucinogen Persisting Perception Disorder' (HPPD), 292.89) and ICD-10 ('Flashbacks', F16.70).

Detailed questions about possible flashback phenomena and spontaneous ASC in the follow-up questionnaire of the present study indicated that none of our subjects fulfilled diagnostic criteria for HPPD in DSM-IV or flashbacks in ICD-10. Furthermore, none of our subjects described visual phenomena reminiscent of the typical symptoms of HPPD mentioned under criterion A of HPPD in DSM-IV (292.89). Our results support the view that HPPD and other troubling perceptual abnormalities rarely occur in a therapeutic or research context, where subjects are carefully screened and monitored and judicious doses of pharmaceutical quality drugs are given (Strassman, 1984; Halpern & Pope, 2003). The clinical relevance of flashback phenomena has been a matter of controversial debate for several decades. Whereas some researchers report virtually no such phenomena in series of hundreds or thousands of cases (Cohen, 1960; McGlothlin & Arnold, 1971), others report incidence rates as high 33% (Moskowitz, 1971) and 77% (Holsten, 1976) among individuals who have taken LSD. A recent review by Halpern and Pope (2003, pp. 116), which analyzed 20 quantitative studies reporting flashback phenomena, concludes that "the data do not permit us to estimate, even crudely, the prevalence of 'strict' HPPD". Halpern and Pope (2003) point out that most of these studies were published before operational diagnostic criteria for HPPD had been established and therefore used a wide variety of methodologies. They further criticize that confounding factors such as recent drug intake, polydrug abuse, preexisting psychiatric disorders, and comorbidity often have been very poorly controlled in the studies.

It should also be noted that the scientific basis for the classification of HPPD in DSM-IV appears to be formed almost exclusively by the research of Abraham and his colleagues (1982; 1983; 1988; 1996; 2001). Since the majority of Abraham's work was focused solely on LSD use, we cannot safely infer that the diagnosis of HPPD has equal implications for all classical hallucinogenic drugs. Interestingly, very few case reports have appeared on flashback-phenomena experienced by individuals who have used psilocybin, DMT, or mescaline exclusively. In fact, Hermle, Kovar, et al. (2008) report in a recent review that there exists only one case report (Espiard, Lecardeur, Abadie, Halbecq, & Dollfus, 2005) where HPPD occurred after hallucinogenic mushroom use.

3.5.3.5 *Prolonged adverse reactions*

The present study supports the notion that prolonged adverse reactions, such as persisting psychosis or depression, are exceedingly rare when psilocybin is administered to well-adjusted subjects in a controlled experimental setting (Strassman, 1984; Abraham and Aldridge, 1993; El-Mallakh et al., 2008). We have found no incidences of prolonged psychotic reactions or precipitations of schizophrenia-spectrum disorders in the 110 subjects studied. However, one of our subjects experienced symptoms of emotional instability, anxiety, and depression, which lasted several weeks and were severe enough for him to seek professional help. A few subjects described occasional mood swings, reactivation of old problems, excessive pensiveness and introversion, and memory and concentration problems in the first few weeks after the drug session. Although these adverse after-effects were generally benign and in all cases resolved after a few weeks, they underline the importance of careful debriefing and follow-up of subjects – especially in the first few days and weeks after drug administration.

3.5.4 *Limitations*

It is important to note that the high degree of safety and tolerability of psilocybin reported in the present study cannot be generalized to situations in which psilocybin is used recreationally or administered under less controlled conditions. It is likely that the careful selection, preparation, and monitoring of subjects as well as the administration of predominantly moderate drug doses have largely contributed to the relatively low occurrence of adverse events in our studies. Our sample might be unrepresentative not only due to exclusion of subjects showing potential risk factors (high emotional lability, history of drug abuse, psychiatric illness, and hereditary risk factors) at screening, but also due to the use of a recruitment method that is prone to self-selection bias. The subjects who volunteered for our studies had prior knowledge that experiments would involve psilocybin administration. Hence, it is likely that individuals who had a positive attitude towards hallucinogenic drugs and who had a personal interest in experiencing drug-induced ASCs were more likely to participate in our experiments. McGlothlin and Arnold (1971) have shown that subjects who are interested in hallucinogenic drug use are susceptible to naturally occurring ASCs, seek to encourage them through both drug and non-drug methods, and have a certain type of personality structure. Indeed, although our subjects were not selected for previous drug-experience, 40% had used a classical hallucinogen at least once in a lifetime prior to the experiments, which is a larger proportion than in the general population.

Although previous experience, positive expectancy, and personality characteristics might have biased subjective drug effects, it may also have contributed to the low occurrence of adverse events. Subjects who expect positive psilocybin effects tend to experience more positive psilocybin effects, whereas anxiety and preoccupation before drug administration not only increases the likelihood of unpleasant experiences, but also the number of somatic complaints (Metzner et al., 1965). Moreover, emotional lability and rigid conventionality – personality traits that tend to be below average in our sample – are positively correlated with DED (Dittrich, 1994). Since safety and tolerability considerations, in our opinion, are more important than methodological rigor, we have not sought to maximize the representativeness of our sample by excluding subjects who had some experience with hallucinogenic drugs – unless they used them on a regular basis. In fact, subjects with a few past experiences were considered ideal because they probably would not have volunteered if they had had significant adverse reactions. This policy has also been proposed by Gouzoulis-Mayfrank, Schneider, et al. (1998).

Apart from the lack of representativeness, our investigation has further limitations. First, cut-off values used to measure the proportion of subjects experiencing very strong subjective drug effects – although similar to specifications used by Pahnke (1969) – are relatively arbitrary, since they neither have been defined a priori nor validated on clear-cut criteria. Second, we have used an investigator-constructed follow-up questionnaire whose reliability and validity is not known.

3.5.5 *Conclusion*

Taken together, our experimental data from 227 psilocybin administrations have demonstrated safety and tolerability not only acutely, but also in the long run. We found

no indication for subsequent drug abuse, persisting perception disorders, prolonged psychosis, or other long-term impairments of functioning in any of our subjects. Acute adverse reactions (so called 'bad' or 'horror trips') occurring in a small proportion of subjects in the two highest dose conditions as well as transient emotional instability lasting a few days or weeks in a small number of subjects remain the biggest concerns in psilocybin administration. However, given that all of these adverse reactions resolved by providing strong interpersonal support and appeared to be positively integrated at the long-term follow-up, 8-16 months after the drug experiments, we conclude that psilocybin administration to healthy, high-functioning, and well-prepared subjects in a responsible clinical or research setting is generally well-tolerated and that future studies using this important research tool are justified.

4

PSYCHOMETRIC EVALUATION OF THE ALTERED STATES OF CONSCIOUSNESS RATING SCALE OAV

4.1 ABSTRACT

The OAV questionnaire has been developed to integrate research on altered states of consciousness ASCs. It measures three primary and one secondary dimensions of ASCs that are hypothesized to be invariant across ASC induction methods. The OAV rating scale has been in use for more than 20 years and applied internationally in a broad range of research fields, yet its factorial structure has never been tested by structural equation modeling techniques and its psychometric properties have never been examined in large samples of experimentally induced ASCs. The present study conducted a psychometric evaluation of the OAV in a sample of psilocybin (n = 327), ketamine (n = 162), and MDMA (n = 102) induced ASCs that was obtained by pooling data from 43 experimental studies. The factorial structure was examined by confirmatory factor analysis, exploratory structural equation modeling (ESEM), hierarchical item clustering (ICLUST), and multiple indicators and multiple causes (MIMIC) modeling. The originally proposed model did not fit the data well even if zero-constraints on non-target factor loadings and residual correlations were relaxed. Furthermore, ICLUST suggested that the "Oceanic Boundlessness" and "Visionary Restructuralization" factors could be combined on a high level of the construct hierarchy. However, because these factors were multidimensional, we extracted and examined 11 new lower order factors. MIMIC modeling indicated that these factors were highly measurement invariant across drugs, settings, questionnaire versions, and sexes. The new factors were also demonstrated to have improved homogeneities, satisfactory reliabilities, discriminant and convergent validities, and to differentiate well among the three drug groups.

4.2 INTRODUCTION

Altered states of consciousness (ASCs) represent a marked deviation in the subjective experience or psychological functioning of a normal individual from her/his usual waking consciousness (Dittrich, 1998). As opposed to psychiatric diseases, ASCs are short-lasting. They are usually self-induced (e.g., by hallucinogenic drugs, meditation, hypnosis), but may also occur spontaneously in everyday life (e.g., hypnagogic states). Although ASCs are by definition different from psychiatric diseases, the study of ASCs in healthy volunteers has a long tradition of generating hypotheses for psychiatric research.

Dittrich's APZ questionnaire and its revised versions, OAV and 5D-ASC, are among the most widely used self-report questionnaires for assessing subjective experiences of ASCs in retrospect. Although originally developed in German, these questionnaires have been translated into many different languages and applied internationally in approximately 70 experimental studies. The majority of these studies have used these

questionnaires to assess ASCs induced by psycho-active drugs, particularly psilocybin (e.g., Griffiths, Richards, McCann, & Jesse, 2006), ketamine (e.g., Northoff et al., 2005), MDMA (e.g., Hasler, Studerus, et al., 2009), and *N*,*N*-dimethyltryptamine (DMT; e.g., Gouzoulis-Mayfrank, Heekeren, Neukirch, et al., 2005), but several studies have also assessed non-pharmacologically induced ASCs, such as ASCs induced by endogenous psychosis (e.g., Gouzoulis-Mayfrank, Habermeyer, et al., 1998), sensory deprivation (Kjellgren, Sundequist, Sundholm, et al., 2004), mind machines (Walach & Käseberg, 1998), and monochrome sounds (Hübner, 2007). The three versions of Dittrich's ASC questionnaires have been successfully applied to differentiate the subjective effects of different ASC induction methods (e.g., Dittrich, von Arx, & Staub, 1985; Dittrich, 1994); to characterize dose-response relationships (Hasler, Grimberg, et al., 2004); and to map first person accounts of ASCs to various neuronal, psychophysiological, and behavioral measures of ASCs, including measures of positron emission tomography (PET; e.g., Vollenweider, Vontobel, Hell, & Leenders, 1999), functional magnetic resonance imaging (e.g., Northoff et al., 2005), and electroencephalography (EEG; e.g., Umbricht, Koller, et al., 2002).

The original version, APZ, contains 158 dichotomous items covering a broad range of phenomena potentially occurring during ASCs. It was originally developed by Dittrich (1975b, 1985, 1996, 1998) in order to test the hypothesis that ASCs – independent of their means of induction – have features in common that can be parsimoniously described on stable (i.e., etiology-independent) major dimensions. Dittrich (1985) reasoned that if this hypothesis could not be falsified for a broad range of ASC induction methods, integration of phenomenological, psychophysiological, and neurobiological research on ASCs would be greatly enhanced. For example, the detection of features that are common/invariant for all ASCs and at the same time differentiate them from normal waking consciousness would help to lay the foundation for a more coherent definition of the term ASC. Furthermore, because these common features of ASCs would remain indicators of the same underlying constructs across different ASC induction methods, ASCs could be characterized and compared by their relative standing on stable etiology-independent latent dimensions. Moreover, establishing such dimensions would eventually lead to an empirical taxonomy of ASCs.

Dittrich (1985) tested his hypothesis in a series of experimental studies, in which healthy volunteers were treated with one of eleven different ASC induction methods (n = 259) or control condition procedures (n = 134). The studied induction methods were divided into four groups: (a) hallucinogens of the first order (i.e., DMT, Psilocybin, and Δ^9-tetrahydrocannabinol); (b) hallucinogens of the second order (i.e., nitrous oxide); (c) sensory deprivation in a broader sense (i.e., perceptual deprivation, hypnagogic states, autogenic training, hypnosis); and (d) sensory overload (i.e., stimuli of high variety). From the 158 items of the APZ, which served as dependent variables, Dittrich (1985) identified 72 items meeting his criteria of etiology-independency. That is, these items had a significant proportion of yes-answers in each group of ASC induction methods and differentiated significantly between treatment and control conditions. By analyzing the correlation matrices of the 72 etiology-independent items using exploratory factor analysis (EFA) and cluster analysis and based on considerations of stability, reliability, and interpretability, Dittrich (1985) determined three oblique primary and one secondary etiology-independent dimensions. The three primary

dimensions were termed Oceanic Boundlessness (OBN), Dread of Ego Dissolution (DED), and Visionary Restructuralization (VRS). The OBN scale basically includes items measuring positively experienced depersonalization and derealization, deeply-felt positive mood, and experiences of unity. High scores on the OBN scale therefore indicate a state similar to mystical experiences as described in the scientific literature on the psychology of religion (e.g., Stace, 1961). The DED scale includes items measuring negatively experienced derealization and depersonalization, cognitive disturbances, catatonic symptoms, paranoia, and loss of thought and body control. High scores on the DED scale therefore indicate a very unpleasant state similar to so called "bad trips" described by drug-users. The VRS scale contains items measuring visual (pseudo)-hallucinations, illusions, auditory-visual synesthesiae, and changes in the meaning of percepts. The secondary scale (G-ASC) consists of the 72 etiology-independent items and can be interpreted as a general measure of consciousness alteration. The validity of the experimental results in the field and the invariance of the factorial structure across different language versions of the APZ was examined and confirmed in a large international study on ASCs (Dittrich, von Arx, & Staub, 1985), in which 1133 subjects from six different countries and four different languages completed the APZ in reference to their most recent ASC that they had experienced within the past 12 months.

Although reliabilities and validities of APZ scales were deemed to be acceptable in the experimental as well as in the field studies, several weaknesses were also recognized. For example, the binary item response format of the APZ was too crude to measure subtle alterations of consciousness. Furthermore, the OBN and VRS dimensions contained a relatively low number of items, and the conceptual breadth of the VRS dimension was considered too narrow. Bodmer (1989) therefore developed a psychometrically improved version called OAV. The abbreviation OAV stands for the German names of the three dimensions OBN, DED, and VRS. Because the OAV was supposed to measure the primary three dimensions of the APZ only, its item pool was primarily derived from 72 etiology-independent items of the APZ. However, the response format was changed from binary to visual analogue, several items were re-worded, some new items were introduced, and some items were completely dropped. The reformulation of items aimed not only at reducing cross-loadings, decreasing ambiguity, and enhancing ease of understanding, but also at widening the conceptual breadth of the OBN and VRS dimensions. Whereas the OBN dimension was changed toward a more complete assessment of mystical experiences by incorporating items that were formulated on the basis of six of the nine categories of mystical experiences proposed by Stace (1961), the VRS dimension was conceptually widened by incorporating items that measure an increase of imagination, associations, and memory retrieval. The re-conceptualization of the VRS dimension was mainly driven by theoretical considerations of Leuner (1962, 1981), who had hypothesized that visual hallucinations are associated with an increased internal stimulus production. The original OAV validation study of Bodmer (1989), which was based on 177 subjects retrospectively describing their most recent ASC, indicated that the questionnaire revision successfully improved several psychometric properties, including item discriminations, simple structure, and scale reliabilities. High correlations of OBN, DED, and VRS scales across the two questionnaire versions suggested that these scales measure similar constructs in both questionnaires. Results obtained by

the APZ and OAV can therefore be compared by transforming the scales through linear equations (see Bodmer, 1999).

Although the dimensional analyses of the APZ and OAV questionnaires had revealed three primary "etiology-independent" dimensions of ASCs, Dittrich's own investigations (Dittrich, 1985), as well as the scientific literature on ASCs, pointed to the existence of further dimensions that are specific to certain ASC-inducing agents. For example, acoustic alterations and hallucinations are a common feature of ASCs induced by certain psychiatric diseases, such as schizophrenia and alcohol withdrawal psychoses, and have been described under conditions of sensory deprivation and hypnagogic states (e.g., Mavromatis, 1987), but seem to be less common in hallucinogen-induced ASCs (Malitz, Wilkens, & Esecover, 1962). In accordance with these findings, only 2 of the 11 APZ items measuring acoustic alterations met criteria of etiology-independency in Dittrich's experimental studies (Dittrich, 1985). Furthermore, clouding of consciousness and reduction of vigilance are characteristic features of hallucinogens of the second order and of sedative drugs, but not of hallucinogens of the first order (Leuner, 1981). Dittrich (1994, 1996) therefore hypothesized that Auditory Alterations (AUA) and Vigilance Reduction (VIR) were two etiology-dependent dimensions, which, in addition to the three primary etiology-independent dimensions, could be reliably and validly measured. In order to measure these dimensions, Dittrich and co-workers constructed and validated the so-called Bewusstseinstrübung und Akustische Halluzinationen (BETA) questionnaire (Dittrich et al., 1989; Schneiter, 1991). Because a dimensional analysis of the Pearson correlation matrix formed from the 39 BETA items and the 49 APZ items comprising the primary three scales indicated that the AUA and VIR dimensions could be differentiated from the OBN, DED, and VRS dimensions and because the reliabilities and validities of the AUA and VIR scales were demonstrated to be acceptable (Schneiter, 1991; Braun, 1997), Dittrich et al. (1999) published an extended version of the OAV, called Five Dimensions of Altered States of Consciousness (5D-ASC), which includes 16 and 12 BETA items measuring the AUA and VIR dimensions, respectively. The 5D-ASC is the latest version of Dittrich's ASC questionnaires. Psychometrically yet untested versions of the 5D-ASC exist in U.S. English (Dittrich et al., 2010), French, Brazilian Portuguese, Arabic, Dutch, and Japanese (A. Dittrich, personal communication, February 14, 2010).

Although Dittrich (1998) concluded that his original hypotheses on ASCs have survived considerable falsification testing not only in experimental but also in field studies and that the APZ questionnaire has become a psychometrically well-validated instrument for the assessment of "etiology-independent" features of ASCs in a "etiology-independent" three-dimensional space, it should be noted that the studies carried out so far have serious methodological limitations, from which only few have been recognized in the existing literature. For instance, the dimensional analyses of the dichotomous APZ items were based on Pearson correlations among the items, which, unlike tetrachoric correlations, can be severely attenuated if the items differ markedly by their difficulties (J. B. Carroll, 1945). This is a serious problem, because it may have led to the extraction of pseudofactors that reflect similar item difficulty rather than similar item content (Kubinger, 2003). Another methodological shortcoming of Dittrich's original investigation is that the stability of the proposed factorial structure across different ASC induction methods and languages has only been examined by descriptive measures of factor pattern similarity derived from the comparison of EFA models, namely, by Tucker's

coefficient of congruence and by Cohen's $\varkappa$. These measures have many recognized problems (Reise et al., 2000). For instance, because they only estimate the similarity of factor loadings, but not the similarity of indicator intercepts and residual variances, they can only provide evidence for the weakest forms of factorial invariance, namely, the so-called "metric invariance" or "weak factorial invariance" in the case of Tucker's coefficient of congruence and "configural invariance" in the case of Cohen's $\varkappa$ (for a description of different levels of factorial invariance, see Meredith, 1993). However, even for the assessment of these weakest forms of factorial invariance, the use of these similarity measures is problematic because their size is affected by various properties of the data (Paunonen, 1997; Barrett, 2006). Thus, it is unclear how large they should be to conclude that factor pattern similarity holds to a reasonable degree. Commonly applied rules of thumb, which were also used in the study of Dittrich (1985), are not only unreliable, they also seem to be much too lenient (Chan et al., 1999). Furthermore, due to his relatively low sample size, Dittrich (1985) assessed the factor pattern similarity across different ASC induction methods only on the level of aggregated items and only across four groups of ASC induction methods. The use of item aggregates, however, is highly problematic when the goal is to represent the dimensionality of the measurement space at the level of individual items (Little et al., 2002).

Another potential bias in Dittrich's original investigation is the use of a specific set of items. Analogous to the dimensions of personality (e.g., the so-called "Big-Five"), broad dimensions of ASCs can only be found by analyzing sets of items that are representative for the domain of interest. Although Dittrich (1985) has originally derived his three primary dimensions of ASCs from a set of 158 APZ items that were selected to be representative for the domain of interest, it is unknown whether the sampling of APZ items was indeed unbiased because his investigation was never repeated in other independent sets of items.

Unfortunately, studies that have re-examined the psychometric properties of Dittrich's ASC rating scales after their first publication are scarce and those that exist were based on very limited sample sizes. Furthermore, because these rating scales were constructed and validated during the early 80s to the mid 90s, dimensional analyses have fully relied on exploratory methods. None of these scales has previously been analyzed by modern latent-variable approaches, such as confirmatory factor analysis (CFA) and structural equation modeling (SEM), which now have become standard methods of psychometric investigations and which are associated with many of the methodological and statistical advances in quantitative psychology in the last two decades (Marsh, Lüdtke, et al., 2010). Because these methods can also overcome many weaknesses of Dittrich's original investigations and more directly assess the validity of Dittrich's hypotheses, studies applying these methods on Dittrich's ASC questionnaires have long been overdue.

Another shortcoming of previous psychometric investigations is that they analyzed only major dimensions and did not explore potential lower order factors or so-called facets, even though applied researchers were not satisfied with the large conceptual breadth of the proposed major dimensions and have constructed their own subscales based on considerations of item content e.g., Liechti, Gamma, and Vollenweider (2001). Addressing these issues is important because Dittrich's questionnaires continue to be widely used, and few validated instruments are available that measure similar subjective experiences. In fact, we are aware of only two instruments that measure

similar experiences and that have gained similar acceptance in applied research, that is, the Phenomenology of Consciousness Inventory (Pekala, Steinberg, & Kumar, 1986) and the Hallucinogen Rating Scale (Strassman et al., 1994).

To overcome the methodological limitations of previous investigations, we performed a psychometric evaluation of the OAV in a relatively large sample of subjects describing experiences of ASCs that were experimentally induced by psilocybin, ketamine, or MDMA. In contrast to previous studies, the factorial structure was explored and tested by using methods of the SEM-framework, including CFA and exploratory structural equation modeling (ESEM; Asparouhov & B. O. Muthén, 2009), and by applying a hierarchical item clustering (ICLUST; Revelle, 1978) procedure that was specifically developed to display the hierarchical structure of a scale. This allowed us to investigate the factorial structure of the OAV not only on a dimensional level, but also on the level of lower order factors or facets. A number of lower order factors were extracted and compared with the original scales. The measurement invariance and population heterogeneity of these lower order factors across different drugs, settings, questionnaire versions, and sexes were examined by MIMIC modeling. The reliabilities were assessed not only by Cronbach's α, but also by various non-standard reliability coefficients for both the original and newly-constructed scales. Furthermore, convergent, discriminant, and known-group validities of these scales were examined. The advantages of the newly constructed subscales, as well as the implications of our results with respect to Dittrich's original hypothesis, are discussed.

4.3 METHODS

4.3.1 *Samples and Data Collection Procedures*

The samples used in the present investigation were obtained by pooling data from 43 experimental studies (including pilot studies) carried out at our research facility between 1992 and 2008 involving psilocybin (115-350 µg/kg po), ketamine (6-12 $\mu g \cdot kg^{-1} \cdot min^{-1}$ iv), or MDMA (1.5-1.7 mg/kg po) administration to healthy volunteers. The studies were part of a research program in which psilocybin, ketamine, and MDMA were used as tools for pharmacological modeling of core symptoms of schizophrenia and for investigating cognitive and perceptual processes (Vollenweider & Geyer, 2001; Geyer & Vollenweider, 2008). All studies were approved by the Ethics Committee of the University Hospital of Psychiatry, Zurich, and the use of psilocybin, ketamine, and MDMA was authorized by the Swiss Federal Office of Public Health, Department of Pharmacology and Narcotics, Berne.

Participants of all studies were recruited through advertisement from the local universities and hospital staff. All subjects signed an informed consent sheet and were carefully screened before admission to the studies. Subjects having personal or family (first-degree relatives) histories of major psychiatric diseases, neurological or substance related disorders, high emotional lability scores (more than two standard deviations above the normative mean in the Freiburg Personality Inventory; Fahrenberg, Hampel, & Selg, 1984), or physical problems (according to a physical examination, electrocardiogram, and clinical-chemical blood test) were excluded. All drug sessions were

Table 6: Descriptive Statistics

Characteristic	Psilocybin $n = 327$	Ketamine $n = 162$	MDMA $n = 102$	Combined $N = 591$	Test statistic[a]
Age ($M \pm SD$)	28.5 ± 6.1	29.5 ± 5.8	26.6 ± 5.1	28.5 ± 5.9	$F(2, 588) = 7.7$, $p < .001$
Gender					$\chi^2 = 31.2$, $p < .001$
Male	57% (187)	80% (130)	76% (78)	67% (395)	
Female	43% (140)	20% (32)	24% (24)	33% (196)	
Education					$\chi^2 = 4.5$, $p = .345$
High school diploma	7% (24)	9% (15)	5% (5)	7% (44)	
University students	57% (187)	55% (89)	67% (68)	58% (344)	
University graduates	35% (116)	36% (58)	28% (29)	34% (203)	
Dose					$\chi^2 = 211$, $p < .001$
Low[b]	22% (72)	0% (0)	0% (0)	12% (72)	
Medium[c]	65% (214)	43% (70)	100% (102)	65% (386)	
High[d]	13% (41)	57% (92)	0% (0)	23% (133)	
Questionnaire version					$\chi^2 = 91.6$, $p < .001$
5D-ASC	69% (227)	26% (42)	38% (39)	52% (308)	
OAV	31% (100)	74% (120)	62% (63)	48% (283)	
Setting					$\chi^2 = 57.7$, $p < .001$
PET	16% (51)	48% (77)	25% (25)	26% (153)	
No PET	84% (276)	52% (85)	75% (77)	74% (438)	

Note. Numbers in parenthesis indicate absolute frequencies. PET = positron emission tomography. [a]based on the comparison between the psilocybin, ketamine and MDMA groups. [b]115 µg/kg psilocybin. [c]215-270 µg/kg psilocybin, 1.5-1.7 mg/kg MDMA, 6 µg · kg^{-1} · min^{-1} ketamine. [d]315 µg/kg psilocybin, 12 µg · kg^{-1} · min^{-1} ketamine.

performed by following safety guidelines that are similar to those recommended by Johnson, Richards, and Griffiths (2008).

In each study, a placebo-controlled within-subject design was used. Depending on the study, subjects received placebo and 1-4 different doses or combinations of psychoactive drugs in 2-5 experimental sessions. Experimental sessions were conducted at least two weeks apart in order to avoid carry-over effects. In the majority of the studies (n = 22), the order of drug administration was randomized and double-blind, but some of the earlier studies as well as most pilot studies (n = 21) were open-label trials. For the present investigation, we only used data from those experimental sessions, in which psilocybin, ketamine, or MDMA was administered alone. Very low dose psilocybin sessions (15-45 µg/kg po) were excluded due to statistically non-significant subjective drug effects (Hasler, Grimberg, et al., 2004). In accordance with these criteria, psilocybin, ketamine, and MDMA were administered in 327, 162, and 102 experimental sessions, respectively. Racemic, (*R*)- and (*S*)-ketamine were administered in 6, 22, and 134 of the ketamine sessions respectively. The total sample consisted of 591 drug sessions. For a detailed description of the sample, see Table 6.

Because some studies involved multiple drug sessions and because some subjects participated in more than one study, the above samples contain non-independent observations. Unfortunately, some of the multivariate statistical procedures used in

the present study rely on the assumption of independency of observations. In order to control for this potential bias, we also analyzed samples that included only one experimental session per subject. For each subject we selected the experimental session that was conducted first. By applying these inclusion criteria, we obtained samples of the following sizes: psilocybin (n = 186), ketamine (n = 109), MDMA (n = 95), and combined drug group (n = 344).

4.3.2 *Measures*

4.3.2.1 *Altered states of consciousness rating scales (OAV and 5D-ASC)*

In each experimental session, subjects were asked to describe the experiences of drug induced ASCs by the German versions of the OAV or 5D-ASC questionnaires. The OAV was used in studies conducted before the year 2000 (n = 27), while the 5D-ASC was used in all later studies (n = 16). Because the 5D-ASC is an extension of the OAV, all 66 OAV items are also fully contained in the 5D-ASC. They also appear in the same order in both questionnaires, but are interspersed by 5D-ASC unique items when presented to the subjects as part of the 5D-ASC. Because the available samples would have been too small to investigate the factorial structures of both questionnaires, items from the 5D-ASC data were combined with the corresponding items of the OAV in the present study. Each OAV item contains a statement describing a specific experience of ASC in the past tense (e.g., "It seemed to me that my environment and I were one"). Subjects were instructed to respond to the described experiences by placing marks on horizontal visual analogue scales (VASs) of 100 millimeters length. The VASs of the OAV are anchored as *no, not more than usual* on the left and as *yes, very much more than usual* on the right. The items are scored by measuring the millimeters from the low end of the scale to the subject's mark (integers from 0-100). Because the low end of the scale indicates a neutral response, the response format of these items can be considered as *strictly unipolar* according to the response format typology of Russell and J. M. Carroll (1999).

In most studies, the OAV and 5D-ASC were completed during or shortly after the drug effects peaked. However, in some studies, these rating scales were completed after the drug effects had worn off or at multiple time points. In the latter case, we only included those measures that were obtained during the peak drug effect. Depending on the study, the pooled OAV and 5D-ASC questionnaires were completed 60-300 min after psilocybin, 25-120 min after ketamine, and 70-160 min after MDMA administration. Subjects were instructed to retrospectively rate their whole experience from the moment of drug intake to the respective measuring time point.

4.3.2.2 *Short Version of the Adjective Word List ("Eigenschaftswörterliste"; EWL-60-S)*

The EWL-60-S (Janke & Debus, 1986) is a German self-report rating scale for the multidimensional assessment of the mental state. It is composed of a list of 60 adjectives (e.g., "anxious", "tired", "sociable"), which can be grouped into 15 subscales each comprising 4 adjectives (see Table 11 for the names of these subscales). The subscales can be further grouped into six domains. Subjects are asked to respond to the adjectives on four-point Likert scales ranging from 0 (*not at all*) to 3 (*very much*). The EWL-60-S is a short version of the original EWL-N and -K questionnaires (Janke & Debus, 1978), which have a very

similar factorial structure but use a dichotomous instead of a Likert-type response format. The EWL-60-S has been found to be well suited to measure short-term changes of mental states induced by psychoactive drugs (e.g., Hasler, Studerus, et al., 2009), psychological stress (Weber et al., 2007), and embodying of emotion (Wiswede, Münte, Krämer, & Rüsseler, 2009). Internal consistencies (Cronbach's α) for the subscales of the EWL-60-S were reported to range between 0.40 and 0.86 in a sample of elderly people (n = 128) and between 0.72 and 0.91 in a student sample (n = 67; Janke & Debus, 1986). The validity of the EWL-60-S has been mostly inferred from the validity studies of the EWL-N and -K, from which most of the EWL-60-S items were taken.

In the present investigation, the EWL-60-S was used to assess the convergent and discriminant validities of the OAV scales. The EWL-60-S was administered in 10 of the 43 pooled studies and in 177 of the 591 analyzed drug sessions. These drug sessions mostly involved the administration of psilocybin (n = 128) and less frequently of (*S*)-ketamine (n = 33) and MDMA (n =16). In cases where the EWL-60-S was administered at multiple time points during one drug session, we used only those measures that were obtained during the peak drug effects. The internal consistencies (Cronbach's α) of the EWL-60-S subscales in our sample were mostly good to excellent and ranged from 0.76 to 0.91.

4.3.2.3 *The State-Trait-Anxiety Inventory – State version (Form X; STAI-S)*

The STAI-S (Spielberger, Gorsuch, & Lushene, 1970; Laux, Glanzmann, Schaffner, & Spielberger, 1981) is a very popular self-report rating scale designed to measure transitory feelings of tension and apprehension, or state anxiety. It contains 10 items describing symptoms of anxiety (e.g., "I feel nervous") and 10 items describing the absence of anxiety (e.g., "I feel calm"). The German translation of the STAI-S has shown excellent internal consistency (average $\alpha \approx .90$) and adequate convergent and discriminant validities with scales of the original EWL questionnaire (Laux et al., 1981). Furthermore, the revised English version of the STAI-S (Form Y) has demonstrated high sensitivity for the detection of stress (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). However, despite these generally positive psychometric properties, the STAI-S has been criticized for its inability to adequately discriminate between symptoms of anxiety and depression (Endler, Cox, Parker, & Bagby, 1992; Grös, Antony, Simms, & McCabe, 2007) and for its lack of unidimensionality. Most of the studies investigating the dimensionality of the STAI reported results indicating that the STAI-S scale could be further divided on the basis of whether the items were keyed in the direction of the presence or absence of anxiety (Vigneau & Cormier, 2008). Consistent with the view that state-anxiety is better accounted for by two unipolar instead of one bipolar construct, the state anxiety present and state anxiety absent scales have been shown to be differentially affected by situational (Kvaal, Laake, & Engedal, 2001) and cultural (Iwata & Higuchi, 2000) variables.

Because the OAV contains subscales tapping symptoms of anxiety as well as the absence of anxiety/well-being, the three STAI-S scales (total scale, anxiety present, and anxiety absent/calmness) were used to assess convergent and discriminant validities of the OAV scales. The STAI-S was concurrently administered with the OAV in 56 of the pooled experimental drug sessions, 45 of which were MDMA and 11 of which were psilocybin sessions. All three subscales showed good internal consistencies in our sample (total scale: α = 0.88, anxiety present: α = 0.82, anxiety absent: α = 0.84).

4.3.3 *Statistical Analysis*

The originally hypothesized factorial structure of the OAV was tested by CFA and ESEM (Asparouhov & B. O. Muthén, 2009; Marsh, B. O. Muthén, et al., 2009) using M*plus* Version 5.2 (L. K. Muthén & B. O. Muthén, 2007). ESEM is a recent statistical development currently only available in M*plus* that integrates many advantages of EFA and CFA by including an EFA measurement model part into a SEM framework. In the present study, we complemented the CFA with ESEM because the imposed simple structure of CFA models, that is, constraining non-target factor loadings to zero, is often inappropriate when analyses are done at the item level and when there are multiple factors, each measured with a reasonable number of items (Marsh, Lüdtke, et al., 2010). Furthermore, ESEM allowed us to perform an EFA while at the same time having full access to all the usual SEM parameters and also taking method effects into account, which may have resulted from items sharing similar wording. Whereas in conventional EFA, method effects can confound the detection of more meaningful factors, they can be controlled in ESEM by allowing correlated residuals (Asparouhov & B. O. Muthén, 2009).

In order to more fully explore the adequacy of the hypothesized three-dimensional solution, the appropriate numbers of factors to extract was examined by means of Cattell's scree test (Cattell, 1966), Horn's parallel-analysis (Horn, 1965), Velicer's minimum average partial (MAP) test (Velicer, 1976), Revelle's very simple structure (VSS) criterion (Revelle & Rocklin, 1979), and Revelle's hierarchical item clustering (ICLUST) algorithm (Revelle, 1978) using functions provided by the nFactors- (Raiche & Magis, 2009) and psych-packages (Revelle, 2009) of the statistical software R (R Development Core Team, 2009).

Because a well fitting simple structure CFA model with clearly defined factors, that is, factors that were measured by at least 3 items and that were conceptually meaningful, was impossible to achieve by the traditional EFA approach and by retaining all 66 OAV items in the solution even when the number of factors was greatly increased, we used cluster analysis as an alternative heuristic for initial CFA model specification. Although rarely used in applied research, a simulation study by Bacon (2001) suggests that cluster analytic approaches to initial model specification are valuable alternatives to the more conventional EFA-related approaches because they may lead to better fitting initial CFA models, which in turn reduces the need for extensive CFA model refinement and consequently the dangers of so-called specification searches.

We applied Revelle's ICLUST procedure, a cluster analytic approach that was specifically developed to cluster questionnaire items and that was recently implemented in the freely available psych-package (Revelle, 2009) of the software R. ICLUST hierarchically clusters items using correlations corrected for attenuation as a proximity measure and the size of the reliability coefficients Cronbach's α and Revelle's β (Revelle, 1979) as stopping rules. A major advantage of ICLUST is that items are only added to clusters if they increase the cluster's internal consistency and factorial homogeneity. Furthermore, as the sequential item-by-item growth of clusters mapped with an accompanying set of homogeneity statistics can be displayed in a hierarchical tree diagram, the ICLUST procedure provides uniquely useful diagnostic and interpretative information not available in conventional approaches of scale construction, such as EFA (Cooksey & Soutar, 2006). For instance, the internal substructure of scales can be directly visualized, and

defensible decisions can be made on whether to form scales on a macro level (higher order scales) and at a more finely grained micro level (lower order scales). Because problematic items usually get merged in a late step of the ICLUST procedure, they can be more easily identified, and they do not obscure the factorial structure as much as in an EFA (for more information, see Cooksey & Soutar, 2006).

An initial simple structure CFA model with correlated latent factors was specified and evaluated on the basis of ICLUST item clusters meeting the following criteria: Satisfactory indexes for internal consistency (Cronbach's $\alpha > 0.8$) and homogeneity (Revelle's $\beta > 0.7$), a minimal cluster size of three items, good interpretability, and conceptual importance. The initial factorial solution was then further refined by dropping items with high cross-loadings.

After having established a well fitting CFA model in the total sample, we used MIMIC modeling to examine population heterogeneity and differential item functioning (DIF) across different drugs, questionnaires, settings, and sexes. Although the multiple-groups CFA approach is more commonly used to examine structural and measurement invariance, we decided to use MIMIC modeling because it requires lower sample sizes and allows the simultaneous evaluation of many different contrast variables (Brown, 2006). Whereas multiple-groups CFA entails the simultaneous analysis of two or more measurement models, MIMIC involves a single CFA model in which latent factor and item indicators are regressed on covariates. A significant direct effect of a covariate on a latent factor indicates that the mean of the latent factors differs across different levels of the covariate (also referred to as population heterogeneity). A significant direct effect of a covariate on an observed item indicator is evidence for measurement non-invariance because it means that the item endorsement is significantly different across different levels of the covariate even though the latent factor is held constant. The occurrence of measurement non-invariance (also referred to as DIF) in MIMIC corresponds to the occurrence of non-invariant item intercepts in multiple-groups CFA and has important consequences for the interpretation of latent factor means. That is, if DIF is present, group comparisons of latent factor means are confounded by group differences in the factor structure and therefore cannot be meaningfully interpreted unless group comparisons are made within the SEM framework, where DIF can be accounted for (Brown, 2006). In the present study, we first examined a MIMIC model in which only the latent factors were regressed on the covariates. Because all direct effects between the covariates and the items were fixed to zero, this constituted the no-DIF model. The latent factors were regressed on the three binary variables female (0 = male, 1 = female), PET (0 = experimental session involved no PET, 1 = experimental session involved PET), and OAV (0 = 5D-ASC, 1 = OAV) and the three-level nominal variable drug. The variable drug was represented in the model as two dummy coded contrast variables using the MDMA group as the reference group. For each of the five binary variables included in the MIMIC model, the minority or focal group contained at least 100 cases. A recent simulation study (Woods, 2009) suggests that focal groups of this size are large enough to produce reasonably powerful and accurate MIMIC results when the sample size is similar to our study. To detect differential functioning (D-F) items we used the so-called "free baseline designated anchor approach" (for applied examples, see Fleishman, Spector, & Altman, 2002; Woods, Oltmanns, & Turkheimer, 2009), which is supposed to have a lower false discovery rate than the more commonly applied stepwise forward procedure (e.g.,

Brown, 2006) and is most similar to well tested item response theory based methods (Woods, 2009). The procedure involved two steps. First, anchor items were identified by regressing one item at a time on the five grouping variables (while constraining all other direct effects to zero) and testing the five regression parameters for significance. Items with no significant regression parameters were defined as DIF-free or anchor items. In the second step, all items not included in the DIF-free subset were tested for DIF by using likelihood ratio (LR) difference tests for nested models. That is, for each studied item, a comparison was made between a full model (all items were allowed to have DIF except for the anchor items) and a more constrained model (all items were allowed to have DIF except for the anchor items and the studied item). If the model fit of the constrained model was significantly worse relative to the full model, it was concluded that the studied item had DIF. As recommended by Woods (2009), p-values of LR difference tests were adjusted by the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) using the p.adjust function in R to control the false discovery rate. After all D-F items were identified, a model was fitted in which – in addition to the latent variables – only D-F items were regressed on the grouping variables. However, to further increase model parsimony, direct effects that were non-significant at $p < 0.05$ or had very low effect sizes (y-standardized regression coefficients < 0.2) were dropped in the final MIMIC model.

Because most OAV items were positively skewed (mean =1.25, range = -0.56 to 4.32) and kurtotic (mean = 1.27, range = 1.64 to 19.23) and because our data set contained non-independent observations, latent factor models (CFA, ESEM, and MIMIC) were fitted by using the Robust Maximum Likelihood (MLR) estimator in combination with the "Complex" option in Mplus. This method produces adjusted standard errors and fit indexes for non-normal and clustered data by means of a sandwich estimator and the Yuan-Bentler T_2* test statistic (Yuan & Bentler, 2000; L. K. Muthén & B. O. Muthén, 2007). Because the χ^2 statistic of the MLR estimator cannot be used for difference tests, the Satorra-Bentler scaled χ^2 difference test (Satorra & Bentler, 2001) was used for the comparison of nested models.

Unfortunately, most OAV items were not only positively skewed and kurtotic, but also showed a strong piling up of values at the lower end of the scale (39% of zero values on average and 71.3% at the most) and a modest piling up of values at the upper end of the scale (6.3% of values on average and 27% at the most; see Supplementary Table S2 for the distributional characteristics of each item). Because parameter estimates produced by MLR can be biased to some degree if strong floor- or ceiling-effects are present, we cross-checked our results by categorizing the OAV items into 5 categories and using the polychoric correlation matrix calculated from categorized variables as input for the latent factor models and the ICLUST procedure. Polychoric correlations estimate the linear associations of latent continuous variables underlying the manifest categorical variables. Because polychoric correlations rest on the assumption that the underlying continuous variables are bivariate normal and that the observed responses are determined by the respondents' thresholds, which are usually converted to the underlying continuous variable by a non-linear transformation, items were categorized such that the degree of normality of the categorized variables was maximized (see Supplementary Table S3 for the distributional characteristics of the categorized items). Although polychoric correlations seem to be robust to modest violations of underlying bivariate normality

(Flora & Curran, 2004), it is generally recommended to test this assumption by the LR-χ^2 statistic and to evaluate model fit by the root mean square error of approximation (RMSEA; Schmukle & Egloff, 2009). According to simulation studies, no serious effects of non-normality on the polychoric correlation estimate are expected to occur as long as RMSEA values are below 0.1 (Jöreskog, 2005). The cut-points that we used in the present study (1, 30, 70, and 90 across all items) produced RMSEA values that slightly exceeded 0.10 in only 2 of 2211 estimated polychoric correlations and therefore should not have resulted in biased parameter estimates. However, it should be noted that polychoric correlations can still be unreliable when estimated on the basis of frequency tables with low expected cell frequencies (Jöreskog, 2005). Because we used 5 categories and equal cut-points across all items, it was unavoidable that some cells of the frequency tables of item pairs became empty. Although a lower number of categories would have reduced the number of empty cells, it would have also increased the information loss resulting from the item categorization. Hence, for the lack of better alternatives, we had to accept this small degree of potential bias.

Another reason why we decided to cross-check our results by using categorical variable methodology was the occurrence of L-shape patterns in bivariate scatter plots of several OAV item pairs describing experience of opposite affective valence. L-shaped bivariate distributions are well known to occur between strictly unipolar items measuring opposite halves of one underlying bipolar continuum because giving a positive response in one item implies giving a neutral response in the other item and vice versa (Schmukle & Egloff, 2009). Due to the opposite skew of such items, the size of their correlation is markedly attenuated when estimated by the product-moment correlation. In fact, it can be shown that the negative correlation between two items measuring mutually exclusive parts of one bipolar continuum theoretically cannot exceed -.47 (Russell & J. M. Carroll, 1999). The only exception to this rule is when the sample is highly polarized (for more details, see Segura & González-Romá, 2003). In contrast to the product-moment correlation, the polychoric correlation can correctly estimate a perfect negative correlation when an L-shape pattern is present. Since most factor analyses in the past have relied on product-moment correlations, the above issue has often led to biased estimates of correlations between latent variables or to the unexpected finding of two separate unipolar dimensions when only one bipolar dimension was hypothesized (see Schmukle & Egloff, 2009). Because the OAV contains strictly unipolar items and because some OAV item pairs of opposite affective valence tended to form L-shape patterns, we hypothesized that the association between these items would become more negative when estimated by polychoric instead of product-moment-correlations and that this would eventually lead to more negative correlations between latent factors for which these items were indicators. Furthermore, we hypothesized that the attenuation of product-moment correlations induced by the L-shape patterns could explain at least in part the unexpected high positive correlation between OAV factors covering experiences of opposite affective valence (e.g., OBN with DED) reported in previous factor analyses (Bodmer, 1989, 1999; Habermeyer, 1999).

In the latent factor models, the polychoric correlation matrix of the categorized variables was used in combination with the robust weighted least squares estimator (WLSMV in M*plus*). The WLSMV provides weighted least square parameter estimates using a diagonal weight matrix and robust standard errors and a mean- and variance

adjusted χ^2 (L. K. Muthén & B. O. Muthén, 2007). Due to good performance in Monte Carlo studies, the WLSMV is currently the method of choice for the analysis of skewed categorical data in small to moderate samples (Flora & Curran, 2004; Brown, 2006).

Because the χ^2 of exact model fit is strongly influenced by sample size, adequacy of fit of the latent factor models was evaluated by Bentler's comparative fit index (CFI), the Tucker-Lewis index (TLI), and the RMSEA. Additionally, the standardized root mean square residual (SRMR) was used for models with continuous outcomes, and the weighted root mean square residual (WRMR) was used for models with categorical outcomes. In line with recommendations of Hu and Bentler (1998), CFI and TLI values close to .95 or greater, RMSEA values close to .06 or below, and SRMR close to .08 or below were considered as indicating reasonably good model fit. For the WRMR, a cut-off value close to 1.0 or below was considered suitable (Yu, 2002).

Because the newly constructed OAV scales did not meet the assumption of essential tau-equivalence (i.e., equality of factor loadings) and because the original OAV scales additionally were non-congeneric, that is, they contained several group factors in addition to a general factor, we did not primarily rely on Cronbach's α for assessing scale reliability. Although α is the most commonly used reliability estimate, it has long been pointed out by several authors that α is not a dependable estimator of scale reliability when the above assumptions are not met and that several alternative reliability estimates exist that obviate the difficulties encountered with the use of α and that can be more easily interpreted (Revelle, 1979; Zinbarg, Revelle, Yovel, & Li, 2005; Green & Yang, 2009). In the present study, scale reliabilities of the newly constructed scales, which had been shown to be congeneric in the CFA, were directly derived within the SEM framework by using an approach described by Raykov (2001, 2009). Point estimates of these reliability estimates, hereinafter referred to as ρ_{SEM}, were supplemented by confidence intervals found by the so-called delta-method (e.g., Raykov & Marcoulides, 2004) to gain ranges of plausible values for the population scale reliabilities. For the original scales not meeting the assumption of unidimensionality, scale reliability was mainly assessed by using McDonalds ω_h and ω_{tot} (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005). Whereas ω_h estimates the amount of variance in a scale attributable to one common factor, also referred to as general factor saturation, ω_{tot} estimates the amount of variance due to all common factors (i.e., group factors *and* general factor). As we had no clear expectations regarding the number of group factors present in the original OAV scales as well as regarding the patterns of the loadings on the common factors, ω_h and ω_{tot} were estimated by performing a higher order EFA analysis using the omega function of the psych-package (Revelle, 2009) in R. This method has shown good performance in a simulation study (Zinbarg, Revelle, & Yovel, 2007). For each scale, the number of group factors to extract was determined by parallel analyses (Horn, 1965). As an alternative estimate of the general factor saturation and as an index of homogeneity, we also computed Revelle's β (Revelle, 1979) using the ICLUST function in the psych-package. Furthermore, Cronbach's α was calculated in order to compare our results with the standard estimate of scale reliability and with the results of older studies. Confidence intervals for α were calculated using the method described by Duhachek and Lacobucci (2004)

Criterion validities of the original and newly constructed OAV scales were evaluated by assessing convergent and discriminant validities, as well as known-group validities.

Convergent and discriminant validities were assessed by correlating the OAV scales with subscales of the EWL-60-S and the STAI-S. Known-group validities were examined by comparing the mean OAV scale scores of the three drug groups.

4.4 RESULTS

In contrast to our hypothesis, associations between items of opposite affective valence with L-shaped bivariate distributions and between latent constructs measuring opposite affective valence did not become more negative when estimated by polychoric instead of product-moment correlations. In fact, polychoric correlations were almost always more positive than product-moment correlations. The difference between the two estimation methods was 0.11 on average. However, because the factorial solution resulting from the analysis of categorized OAV items did not markedly differ from the factorial solution of the continuous items, except that the correlations between the latent variables were generally larger, we only report results from those statistical analyses that treated the OAV items as continuous variables. Results from analyses based on categorized variables are available upon request from the first author.

4.4.1 *Fit of the Originally Hypothesized Model*

Table 7 provides fit indexes from a series of latent factor models testing Bodmer's originally hypothesized factorial structure of the OAV. When modeled as simple structure CFA with no correlated residuals and unconstrained latent factor covariances, Bodmer's factorial structure did not fit the data well. Although the parsimony-adjusted RMSEA and the absolute fit index SRMR were only slightly above the recommended cutoffs, comparative fit indexes (i.e., CFI and TLI) were clearly unacceptable. Large modification indexes for the residual covariances between the items 8, 13, 20, and 25 and between the items 14 and 51, as well as similar wordings within these two item clusters, suggested that item covariances within these two item clusters might be explained in part by shared method effects. We therefore specified a less constrained CFA model in which residual covariances between the items 8, 13, 20, and 25 and between the items 14 and 51 were allowed to freely co-vary. Although this model fitted significantly better than the original model according to the scaled χ^2-difference test, comparative fit indexes were still clearly unacceptable. Because the imposed simple structure of a standard CFA model is often unnecessarily restrictive (Marsh, Lüdtke, et al., 2010), we next tested geomin- and quartimin-rotated 3-Factor-ESEMs with and without method effects. As expected, the free estimation of cross-loadings significantly improved model fit. Because all available fit indexes, including the parsimony-adjusted RMSEA, improved, the increase in model fit was not primarily achieved at the expense of increased model complexity. However, even when method effects were taken into account, overall model fit was relatively poor as the CFI and TLI values were still far below their recommended cutoffs. An inspection of the item loadings of the geomin-rotated ESEM without method effects revealed that 59 of the 66 items (89.4%) had their highest loading on the hypothesized factors. Six VRS items (# 17, 18, 37, 40, 52, and 64) describing experiences of changed meaning of percepts, facilitated recollection, and insightfulness loaded highest on the OBN factor and one VRS item (# 58) describing experiences of macropsia

Table 7: Confirmatory Factor Analysis Model Fit Results

	df	MLR χ^2	CFI	TLI	RMSEA	SRMR	AIC	Δdf	$\Delta\chi^2$	*p*
Bodmer's original structure										
CFA (simple structure)	2076	6661.7	.709	.699	.061	.094	363480			
CFA with method effects	2069	6190.6	.738	.729	.058	.089	362819	7	393	< .001
ESEM	1950	5469.9	.777	.754	.055	.050	150706	119	670	< .001
ESEM with method effects	1943	5235.3	.791	.769	.054	.054	150103	7	37	< .001
Bifactor model	2013	5487.6	.779	.765	.054	.086	150545	−70	−223	
Bifactor model with method eff.	2006	5116.9	.803	.789	.051	.095	150061	7	942	< .001
OBN factor alone	324	1643.7	.785	.767	.083	.070	64550			
DED factor alone	189	612.5	.834	.815	.062	.070	42944			
VRS factor alone	135	1199.5	.689	.648	.116	.090	45175			
Model revision										
Initial ICLUST solution	979	2016.0	.907	.897	.042	.059	255679			
Final model	764	1430.8	.929	.921	.038	.052	94687			
MIMIC models										
Final model: MIMIC without DIF	919	1780.6	.918	.904	.040	.050	97624			
Final model: MIMIC with DIF	910	1668.3	.928	.915	.038	.048	97489	9	120	< .001

Note. MLR = maximum-likelihood-robust estimator; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike's information criterion; $\Delta\chi^2$ = Satorra-Bentler scaled χ^2 difference; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; OBN = oceanic boundlessness; DED = dread of ego dissolution; VRS = visionary restructuralization; ICLUST = hierarchical item-clustering; MIMIC = multiple indicators multiple causes; DIF = differential item functioning.

and micropsia loaded highest on the DED factor. Although all items had at least one significant main factor loading of at least modest size (> 0.3), 28 items demonstrated also significant cross-loadings. The geomin-rotated ESEM that included method effects showed a considerably different pattern of factor loadings. In this model, only 54.5% of the items were correctly distributed to their hypothesized factors. Whereas the OBN and VRS factors essentially collapsed into one large first factor, the DED factor was divided into one factor tapping experiences of anxiety and another factor tapping experiences of impaired control and cognition. Supplementary Table S4 and S5 show the hypothesized and empirical item distributions resulting from the geomin- rotated 3-Factor-ESEMs with and without method effects, respectively. Quartimin-rotated ESEMs only marginally differed from their geomin-rotated counterparts.

Because previous EFAs of the APZ and OAV had revealed large first eigenvalues relative to later eigenvalues and because the existence of a general factor has been hypothesized for both the APZ and OAV (Dittrich, 1985; Bodmer, 1989), we also tested a bi-factor model in which all items were allowed to load on a general factor in addition to their specific group factor. Although this model fit better than all previously tested models, comparative fit indexes were still clearly unacceptable. In order to examine the homogeneity of the hypothesized factors, we also modeled each factor separately. The results indicated that none of the three hypothesized factors can be considered unidimensional and that VRS is the most heterogeneous factor.

4.4.2 *The Optimal Number of Factors to Extract*

Although the OAV questionnaire was specifically designed to measure three dimensions of ASCs, none of the methods that we used to determine the optimal number of factors to extract indicated a three-dimensional solution. Parallel analysis, which is considered as one of the most effective and accurate methods for determining the number of factors to retain (Velicer, Eaton, & Fava, 2000), suggested 5 and 13 factors, depending on whether the analysis was based on principal component (PA-PCA) or principal factor (PA-PFA) eigenvalues, respectively. In case of the PA-PFA, the number of factors was reduced to 11, when the observed eigenvalues were compared with the 95th percentiles instead of the means of the eigenvalues generated from random data. Although in a recent Monte Carlo study, PA-PFA outperformed PA-PCA under conditions similar to our study (presence of correlated factors or strong general factor as well as group factors; Crawford et al., 2010), the scree test supported the results of the PA-PCA by also suggesting a five-factorial solution. However, the MAP test indicated seven factors to retain, while the VSS criterion for complexity one and two favored one- and two-factorial solutions, respectively. Furthermore, the ICLUST algorithm, which clusters scales as long as the homogeneity and internal consistency of the higher level scale is greater than that of either subcomponent, did not stop until two clusters were left. One of these two item clusters comprised all DED items, while the other comprised all OBN and VRS items. Finally, by testing the fit of ESEMs with a varying number of factors, it was determined that at least 11 factors were necessary to achieve acceptable overall model fit. The optimal numbers of factors obtained by the methods discussed above are summarized in Supplementary Table S6.

4.4.3 *Construction of New OAV Scales*

Although ESEMs with 11 or more factors fit reasonably well, they did not serve well as a basis for initial CFA model specification because they contained several poorly defined factors and a relatively large number of items with significant cross-loadings (see Supplementary Table S7 for the loading matrices of ESEMs with an increasing number of factors starting with the originally hypothesized three-factor solution). Instead of dropping multidimensional items step by step and thereby using CFA in an exploratory fashion, which is generally not recommended because it can lead to problematic specification searches (Brown, 2006), we inspected the tree diagram produced by ICLUST to directly derive homogeneous and reliable subscales. By applying the criteria defined in the method section, 11 item clusters formed from 47 of the 66 original items were detected and used for initial CFA model specification. The ICLUST tree diagram and the item clusters that were used for the initial CFA model are shown in Supplementary Fig. S1 (for the ICLUST tree diagram based on the categorized variables see Supplementary Fig. S2). As the model fit of this initial CFA model was not sufficient according to the CFI and TLI indexes (see Table 7), we tried to improve model fit by dropping items showing large modification indexes for cross-loadings and ambiguous item wordings. The model revision led to a final model that still contained the same number of factors, but a slightly lower number of items (42 instead of 47). Because the dropped items (# 12, 39, 41, 48, and 54) had been mostly assigned to different factors,

Table 8: Correlations Between Latent Factors and Observed Correlations of the OAV

Scale	Experience of unity	Spiritual experience	Blissful state	Insightfulness	Disembodiment	Impaired control and cognition	Anxiety	Complex imagery	Elementary imagery	Audio-visual synesthesiae	Changed meaning of percepts
New scales											
Experience of unity	—	.82	.72	.78	.74	.50	.28	.61	.41	.54	.65
Spiritual experience	.69	—	.67	.80	.53	.27	.24	.53	.37	.42	.47
Blissful state	.62	.53	—	.63	.39	.11	−.07	.45	.25	.34	.46
Insightfulness	.63	.61	.49	—	.49	.34	.30	.61	.45	.47	.67
Disembodiment	.64	.45	.35	.39	—	.64	.37	.49	.33	.44	.48
Impaired control and cognition	.43	.24	.11	.27	.53	—	.70	.43	.36	.35	.62
Anxiety	.25	.22	−.06	.24	.32	.63	—	.23	.26	.20	.39
Complex imagery	.52	.44	.39	.48	.40	.36	.20	—	.73	.65	.53
Elementary imagery	.37	.34	.25	.36	.29	.33	.27	.62	—	.70	.41
Audio-visual synesthesiae	.49	.38	.32	.39	.39	.31	.18	.56	.62	—	.44
Changed meaning of percepts	.55	.39	.40	.53	.38	.50	.32	.42	.35	.38	—
Original scales											
Altered state of consciousness	.83	.66	.57	.68	.70	.69	.51	.70	.62	.64	.72
Oceanic boundlessness	.92	.76	.76	.71	.72	.46	.25	.58	.43	.52	.62
Dread of ego dissolution	.39	.26	.03	.30	.50	.91	.86	.34	.35	.29	.48
Visionary restructuralization	.62	.50	.46	.61	.47	.49	.32	.82	.79	.78	.69

Note. Correlations between latent factors are above the diagonal. Correlations based on scale sum scores are below the diagonal. All correlations were statistically significant at $p < .001$ except for blissful state with dread of ego dissolution, anxiety, and impaired control and cognition.

the model revision did not lead to a major change in the interpretation of any factor. Figure 5 shows the factorial structure of this final model, including the names that we gave to these 11 factors and the fully standardized loadings and error variances. The correlations between the latent factors, as well as their associations with the original OAV scales, are shown in Table 8. Although the CFI and TLI of this final CFA model were still slightly below the recommend cutoffs, the RMSEA and the SRMR indicated excellent model fit (see Table 7).

To assure that the parameters of the final model were estimated with sufficient accuracy and that statistical power was high enough to detect significant effects, we performed a Monte Carlo analysis in M*plus* as described by L. K. Muthén and B. O. Muthén (2002). The parameter values from the final model were used as the population parameter values, the sample size was set to 591, and the model estimation was repeated 10,000 times. The Monte Carlo analysis demonstrated that model parameters and their

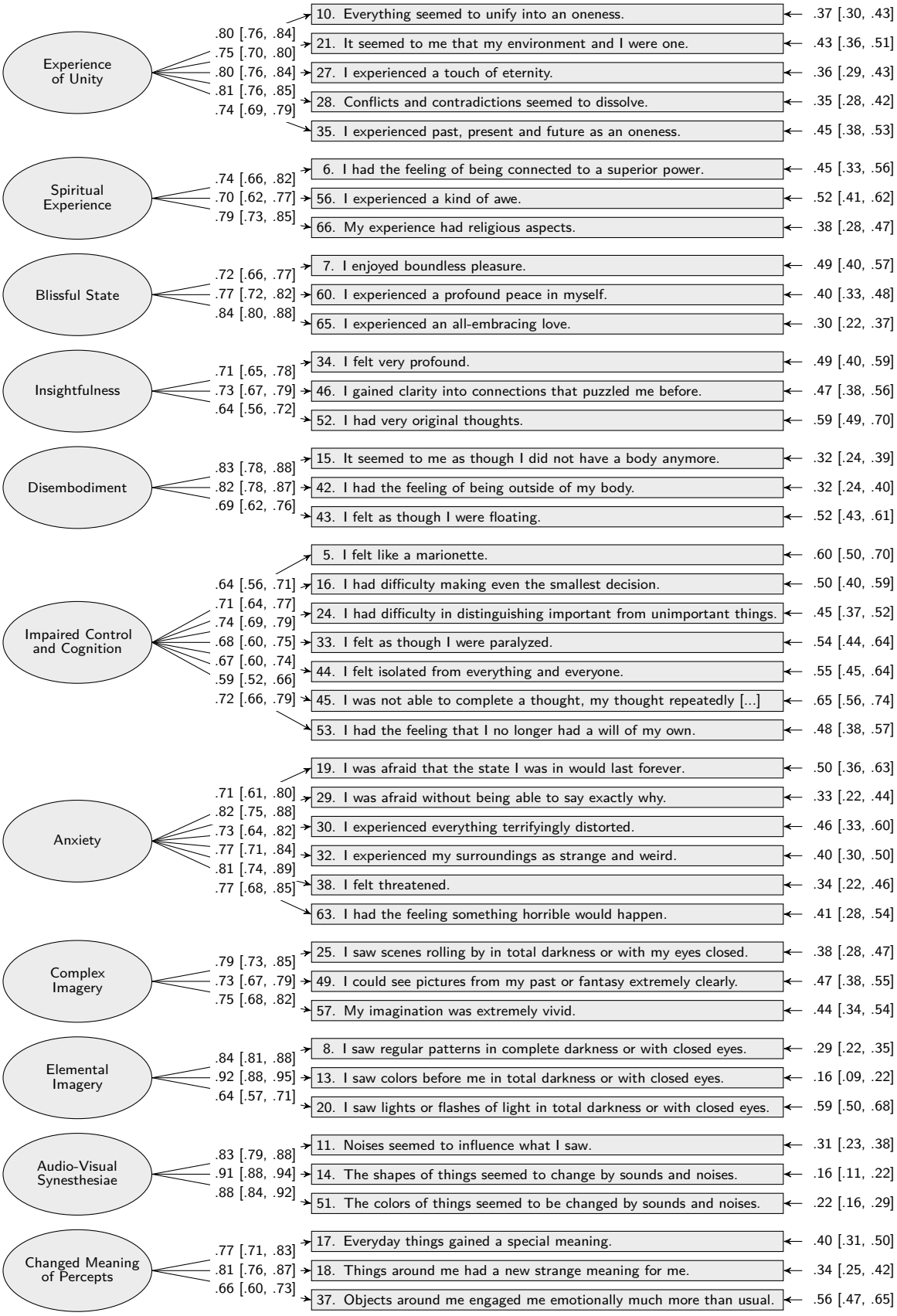

Figure 5: Final confirmatory factor analysis model with completely standardized loadings and error variances. Numbers in brackets are the 95% confidence intervals for the estimates. Covariances between factors were freely estimated and are shown in Table 8.

standard errors were relatively stable and powerful. Specifically, all parameter estimates and their standard errors had bias less than 5%, coverage of all parameter estimates was within the recommended range of 0.91-0.98, and power was higher than 0.8 for all parameter estimates, except for two factor covariances of small effect sizes.

4.4.4 *MIMIC Modeling*

The no-DIF MIMIC model showed only slightly reduced global model fit relative to the final CFA model (see Table 7). This suggests that the associations between the covariates and the items were mostly well explained by the indirect effects going through the latent factors. However, by applying the full baseline designated anchor approach to DIF detection, as outlined in the method section, six D-F items were identified (item # 18, 25, 27, 30, 32, and 33). The estimation of a MIMIC model that included direct effects from each covariate to each D-F item ($5 \times 6 = 30$ direct effects) revealed that nine direct effects were statistically significant at $p < 0.05$ and of at least small to moderate effect size (y-standardized regression coefficient > 0.2). The final MIMIC model, which accounted for DIF by allowing these 9 direct effects to be freely estimated, fitted significantly better than the no-DIF model (see Table 7), and showed reasonably good global model fit. As with the final CFA model, we performed a Monte Carlo analysis to assure that the parameters of the MIMIC model with DIF adjustment were estimated with sufficient power and accuracy. The analysis confirmed that the parameter estimates and their standard errors were relatively stable and powerful.

From the nine direct effects, six effects (those on item # 18, 25, 27, 30, 32, and 33) were due to measurement non-invariance between the MDMA and ketamine groups. Measurement non-invariance between males and females, between the OAV and 5D-ASC questionnaires, and between the MDMA and psilocybin groups was each accounted for by one direct effect (those on item # 18, 30, and 25, respectively). Whereas the direct effects of the two drug contrasts were well explainable by specific effects of psilocybin and ketamine, the direct effects of the gender and questionnaire version covariates were more difficult to interpret. However, since all estimated direct effects were of only small to moderate effect sizes (all y-standardized regression coefficients were between 0.2 and 0.5), these effects must be interpreted cautiously. Furthermore, because the effects of the covariates on the latent factors in the final MIMIC model with DIF adjustment were not substantially different from those of the no-DIF MIMIC model, the impact of the nine direct effects on the estimated group differences in latent factor means can be considered low. In fact, none of these estimated group differences changed statistical significance as a consequence of controlling for DIF.

The effects of the grouping variables on the latent factors in the DIF-adjusted final MIMIC model are shown in Table 9. Compared to MDMA, psilocybin had the most pronounced effect on scales measuring visual alterations (i.e., elementary and complex imagery, audio-visual synesthesiae, and changed meaning of percepts), but also facilitated insights and spiritual experiences and slightly increased anxiety. Ketamine, on the other hand, most strongly reduced blissfulness, increased disembodiment, and impaired control and cognition. Although the effects were less pronounced than those of psilocybin, ketamine also induced visual alterations, most notably elementary imagery, and facilitated spiritual experiences. Furthermore, ketamine slightly increased anxiety

Table 9: Y-Standardized Regression Coefficients [and 95% Confidence Intervals] of the Final MIMIC Model With DIF

Factor	PET	Female	OAV	Psilocybin	Ketamine
Experience of unity	**0.27**	**0.25**	0.12	0.13	0.21
	[0.06, 0.47]	[0.03, 0.47]	[-0.09, 0.34]	[-0.14, 0.39]	[-0.07, 0.48]
Spiritual experience	**0.26**	0.09	**0.27**	**0.43**	**0.31**
	[0.04, 0.49]	[-0.18, 0.36]	[0.04, 0.50]	[0.16, 0.70]	[0.05, 0.58]
Blissful state	0.20	0.06	**0.28**	-0.27	**-0.79**
	[-0.02, 0.41]	[-0.16, 0.27]	[0.08, 0.48]	[-0.54, 0.01]	[-1.06, -0.52]
Insightfulness	**0.30**	-0.17	**0.32**	**0.60**	0.24
	[0.08, 0.52]	[-0.40, 0.06]	[0.07, 0.56]	[0.33, 0.88]	[-0.04, 0.52]
Disembodiment	**0.53**	**0.25**	0.03	0.14	**0.75**
	[0.33, 0.72]	[0.04, 0.46]	[-0.17, 0.23]	[-0.09, 0.37]	[0.51, 0.99]
Impaired control and cognition	**0.31**	**0.36**	-0.13	**0.26**	**0.66**
	[0.10, 0.52]	[0.11, 0.60]	[-0.34, 0.08]	[0.02, 0.49]	[0.40, 0.91]
Anxiety	**0.31**	0.05	0.04	**0.28**	**0.31**
	[0.09, 0.53]	[-0.15, 0.25]	[-0.15, 0.22]	[0.10, 0.46]	[0.08, 0.54]
Complex imagery	**0.45**	0.15	0.02	**0.80**	**0.41**
	[0.24, 0.66]	[-0.08, 0.38]	[-0.20, 0.23]	[0.54, 1.05]	[0.13, 0.69]
Elementary imagery	**0.36**	0.03	0.08	**1.44**	**0.73**
	[0.19, 0.53]	[-0.15, 0.21]	[-0.10, 0.26]	[1.28, 1.61]	[0.53, 0.93]
Audio-visual synesthesiae	-0.04	0.05	0.19	**0.87**	**0.55**
	[-0.23, 0.14]	[-0.17, 0.27]	[-0.02, 0.39]	[0.68, 1.06]	[0.34, 0.75]
Changed meaning of percepts	**0.28**	-0.06	**0.33**	**0.44**	-0.17
	[0.07, 0.49]	[-0.32, 0.21]	[0.11, 0.55]	[0.17, 0.71]	[-0.47, 0.13]

Note. Significant regression coefficients ($p < .05$) are in boldface. By convention, y-standardized regression coefficients of dummy coded variables of the sizes 0.2, 0.5, and 0.8 indicate small, medium, and large effect sizes, respectively. MIMIC = multiple indicators multiple causes; DIF = differential item functioning; PET = positron emission tomography (0 = no PET, 1 = PET); Female (0 = male, 1 = female); OAV = version of the altered state of consciousness rating scale (0 = 5D-ASC, 1 = OAV); Psilocybin (0 = 1.5-1.7 mg/kg MDMA or 6-12 $\mu g \cdot kg^{-1} \cdot min^{-1}$ ketamine, 1 = 115-315 µg/kg psilocybin); Ketamine (0 = 1.5-1.7 mg/kg MDMA or 115-315 µg/kg psilocybin, 1 = 6-12 $\mu g \cdot kg^{-1} \cdot min^{-1}$ ketamine).

compared to MDMA. Averaged over all drugs, females reported more impairment in control and cognition and slightly more/stronger experiences of disembodiment and unity than males. Relative to the 5D-ASC, the OAV questionnaire measured increased changed meaning of percepts, insightfulness, blissful state, and spiritual experiences. Although the different questionnaire lengths and the way items were embedded might have contributed to these differences, it is more plausible that the questionnaire effects were confounded by different drug doses. The average psilocybin doses administered in experiments using the OAV and 5D-ASC were 212 and 251 µg/kg, respectively, whereas the average doses of ketamine were 6 and 12 $\mu g \cdot kg^{-1} \cdot min^{-1}$, respectively.

When drug sessions involved PET measurements, subjects generally experienced stronger subjective drug effects. All scale scores were increased except for the audio-visual synesthesiae and the blissful state scales. The most pronounced effects were observed with respect to visual alterations and disembodiment. The effects of the PET

Table 10: Reliabilities of the OAV Scale Scores

Scale	Items	Cronbach's α	Revelle's β	ρ_{SEM}	ω_h	ω_{tot}
Original scales						
Altered state of consciousness	66	.96 [.96, .97]	.61		.65	.97
Oceanic boundlessness	27	.95 [.94, .96]	.71		.74	.96
Dread of ego dissolution	21	.93 [.92, .94]	.74		.74	.94
Visionary restructuralization	18	.91 [.90, .92]	.73		.70	.93
New scales						
Experience of unity	5	.88 [.87, .90]	.86	.88 [.87, .90]		
Spiritual experience	3	.77 [.74, .81]	.73	.78 [.73, .83]		
Blissful state	3	.82 [.79, .84]	.79	.82 [.79, .85]		
Insightfulness	3	.73 [.69, .77]	.69	.74 [.69, .79]		
Disembodiment	3	.82 [.80, .85]	.77	.82 [.79, .86]		
Impaired control and cognition	7	.85 [.84, .87]	.80	.86 [.83, .88]		
Anxiety	6	.89 [.88, .90]	.83	.89 [.87, .92]		
Complex imagery	3	.80 [.77, .83]	.77	.80 [.77, .83]		
Elementary imagery	3	.84 [.81, .86]	.73	.86 [.83, .88]		
Audio-visual synesthesiae	3	.91 [.89, .92]	.89	.91 [.89, .93]		
Changed meaning of percepts	3	.79 [.77, .82]	.75	.80 [.76, .84]		

Note. Numbers in brackets indicate 95% confidence intervals.

setting might be explained in part by the fact that subjects had more time to concentrate on their experiences when drug sessions took place at the PET center. Specifically, subjects did not have to perform tasks during PET measurements, they were mostly lying in a comfortable horizontal position, and they could have their eyes closed most of the time. However, similar to the questionnaire variable, the setting variable might have been confounded by the effects of different drug doses. The average psilocybin doses administered at the PET center and at the laboratory were 219 and 254 μg/kg, respectively, whereas the average doses of ketamine were 0.79 and 0.87 $\mu g \cdot kg^{-1} \cdot min^{-1}$, respectively.

Although we have controlled the effects of different drug doses in a separate MIMIC model in which we included more dummy variables for different drug groups (i.e., dummy variables for low dose psilocybin, low medium dose psilocybin, high medium dose psilocybin, high dose psilocybin, medium dose ketamine, and high dose ketamine with MDMA as the reference group), we decided to provide the results of this analysis as supplementary material only (see Supplementary Table S8) because a Monte Carlo analysis indicated that the complexity of this model was too high for the size of our sample. Nevertheless, the MIMIC model that included these dose predictors suggested the effects of the PET setting were only slightly confounded by different drug doses. Although the effect sizes were slightly reduced, none of the effects of the setting variable changed statistical significance.

4.4.5 *Reliability Assessment*

The results of the reliability assessment of the original and new OAV scales are shown in Table 10. Because the original scales were demonstrated to be multidimensional in the CFA, it was expected that Cronbach's α would be a biased reliability index for these scales. Indeed, a comparison of α with alternative indexes of reliability revealed that α grossly overestimated reliability, when reliability is defined as the proportion of variance in a scale that is due to one common factor (McDonald's ω_h) and slightly underestimated reliability, when reliability is defined as the proportion of variance due to all common factors (McDonald's ω_{tot}). However, even though variance explained by group factors (i.e., factors that are related to a subset of items within a scale) contributed considerably to the very high α coefficients in the original scales, it should be noted that these scales showed relatively large general factor saturations. Specifically, Revelle's β and McDonald's ω_h were larger than 0.7 for the OBN, DED, and VRS scales and still exceeded 0.6 for the total scale (i.e., G-ASC). Thus, although the original scales are not unidimensional, the general factors (i.e., factors that are common to all items in a scale) clearly dominated these scales because they explained more than 70% of the variance in the OBN, DED, and VRS scales and more than 60% in the total scale. Because these values exceeded the recommended minimum threshold of Revelle (1979), who suggested that the amount of variance explained by the general factor should be at least 50%, and in case of the OBN, DED, and VRS scales even exceeded the more stringent recommendations of Rossiter (2002), who suggested aiming for a coefficient β of 0.7, the calculation of sum scores from these scales, including the G-ASC scale that includes all 66 items, could be justified.

When applied to the new OAV scales, Cronbach's α was a less biased estimator of scale reliability. This was expected, as the new OAV scales had been shown to be unidimensional in the CFA. However, because α underestimates reliability when the items of a scale are not tau-equivalent (i.e., have unequal factor loadings) and because the assumption of tau-equivalence was not met by most of the new OAV scales, ρ_{SEM} was slightly higher than α in these scales. When estimated by ρ_{SEM}, most of the new OAV scales showed good reliabilities. Only two of the 11 scales (insightfulness and spiritual experience) had reliabilities smaller than 0.8. However, both of these scales consisted of only three items, and their reliabilities were still above 0.7, which indicates modest reliability (Nunnally & Bernstein, 1994). By comparing ρ_{SEM} of the new scales with ω_h and ω_{tot} of the old scales it can be seen that, although the new OAV scales have lower reliabilities than the old OAV scales when reliability is defined as the amount of variance in a scale that is due to all common factors, they contain a larger proportion of variance attributable to one common factor. Hence, the new scales are more homogeneous than the old scales. This was also confirmed by the values of coefficient β, which generally were higher for the new scales than for the old OAV scales.

4.4.6 *Validity Assessment*

Pearson correlations between the OAV scales and the 15 subscales of the EWL-60-S computed from the raw sum scores are presented in Table 11. The directions and sizes of the correlations between OAV and EWL-60-S subscales covering similar and dissimilar

Table 11: Correlations Between the OAV and the Adjective Mood Rating Scale (EWL60-S)

	Efficiency-activation	Concentration	Inactivation	Tiredness	Dazed state	Extroversion	Introversion	Self-confidence	Heightened mood	Emotional excitation	Sensitivity	Aggression-anger	Apprehension-anxiety	Depressiveness	Dreaminess
Original scales															
Altered state of consciousness	.10	−.14	.02	−.05	.09	−.04	.09	−.03	.10	.28***	.29***	.01	.14	.07	.24**
Oceanic boundlessness	.18*	.01	−.05	−.12	.02	.06	.03	.14	.23**	.23**	.23**	−.02	.07	.01	.23**
Dread of ego dissolution	−.01	−.24**	.09	.07	.10	−.16*	.16*	−.27***	−.16*	.27***	.23**	.09	.27***	.19*	.10
Visionary restructuralization	.03	−.15*	.03	−.05	.12	−.02	.05	.03	.14	.18*	.25***	−.04	.00	−.03	.27***
New scales															
Experience of unity	.13	.00	−.03	−.13	.01	.04	.04	.11	.21**	.21**	.21**	−.04	.08	.05	.28***
Spiritual experience	.14	.05	−.01	−.08	−.06	−.03	.00	.09	.10	.14	.21**	.01	.09	.01	.10
Blissful state	.19*	.11	−.13	−.12	−.08	.18*	−.04	.27***	.39***	.12	.17*	−.01	−.04	−.08	.18*
Insightfulness	.19*	−.04	−.06	−.09	−.04	.02	−.05	.09	.14	.17*	.17*	−.02	.04	.00	.21**
Disembodiment	.13	−.04	−.04	−.07	.06	−.04	.03	−.08	.04	.18*	.10	−.05	.06	.02	.11
Impaired control and cognition	−.01	−.24**	.09	.03	.17*	−.12	.17*	−.15*	.00	.21**	.21**	.02	.17*	.08	.21**
Anxiety	−.03	−.19*	.09	.11	.02	−.15*	.13	−.29***	−.25**	.24**	.19*	.15*	.32***	.25***	−.03
Complex imagery	.05	−.11	.04	−.01	.11	.02	−.03	.11	.12	.13	.17*	−.03	−.04	−.06	.28***
Elementary imagery	−.10	−.18*	.15	.04	.17*	−.12	.14	−.02	.09	.09	.17*	−.01	.01	−.03	.15*
Audio-visual synesthesiae	.02	.07	.02	−.03	.06	.06	.12	.09	.17*	.11	.32***	.07	.03	.06	.17*
Changed meaning of percepts	.05	−.24**	−.04	−.14	.04	−.04	.00	−.04	.05	.21**	.16*	−.12	.01	−.04	.20**

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

content supported the convergent and discriminant validities of the new OAV scales. For instance, the new OAV anxiety scale correlated highest with the apprehension-anxiety scale of the EWL-60-S, impaired control and cognition correlated highest with concentration, audio-visual synesthesiae correlated highest with sensitivity, vivid imagery correlated highest with dreaminess, and blissful state correlated highest with heightened mood. Compared to the old OAV scales, the new OAV scales tended to correlate higher with scales measuring similar experiences. For example, although the OBN scale correlated highest with the EWL-60-S subscale that was hypothesized to cover the most similar content (i.e., the heightened mood scale), this correlation (r = .27) was considerably lower than the correlation between the more specific blissful state scale and the heightened mood scale (r = .37).

Pearson correlations between the OAV and the STAI-S scales showed that the STAI-S total scale was significantly associated with DED (r = .59, p < .001), anxiety (r = .54, p < .001), and impaired control and cognition (r = .45, p < .001). The STAI-S anxiety present scale correlated significantly with DED (r = .60, p < .001), G-ASC (r = .38, p < .001), impaired control and cognition (r = .52, p < .001), anxiety (r = .51, p < .001), and changed meaning of percepts (r = .33, p = .012), whereas the STAI-S anxiety absent scale correlated significantly with DED (r = .45, p < .001), anxiety (r = .45, p < .001), blissful state (r = -.44, p < .001), and impaired control and cognition (r = .30, p = .024). Although these correlations further support the construct validities of the OAV scales, it should be noted that these correlations were calculated on the basis of a relatively small sample of 56 experimental sessions, primarily involving MDMA administration, which is known to rarely induce anxiety or even has anxiolytic effects (Liechti & Vollenweider, 2000).

Figure 6 displays the mean scores of the new and original OAV scales in the three different drug groups. As can be seen from the plot, the new OAV scales differentiated well among the three drug groups and provided considerably more information on the specific effects of MDMA, ketamine, and psilocybin than the original scales.

4.5 DISCUSSION

This study examined the factorial structure of the OAV questionnaire in a sample of drug induced ASCs by using SEM methodology. The results of this study do not support the three dimensional structure originally proposed by the authors of the OAV (Bodmer, 1989; Dittrich, 1994). The original model provided a poor fit to the data not only when cross-loadings and residual correlations were fixed to zero (simple structure CFA), but also when cross-loadings were freely estimated and residuals of items with similar wording were allowed to freely co-vary (ESEM with method effects) or when an additional general factor was specified (bifactor model).

Although none of the three originally hypothesized OAV factors met criteria of unidimensionality, the results of this study suggest that the VRS factor is the biggest source of misfit. The VRS factor provided not only the worst fit to the data when the three factors were tested separately by one-factor CFAs, it also contained the highest number of items having "wrong" salient factor loadings in the ESEM (six of the seven mis-assigned items were VRS items) and had the lowest general factor saturation. The finding that VRS is the most heterogeneous factor is in agreement with both of the two other studies that have re-examined the factorial structure of the OAV after its first

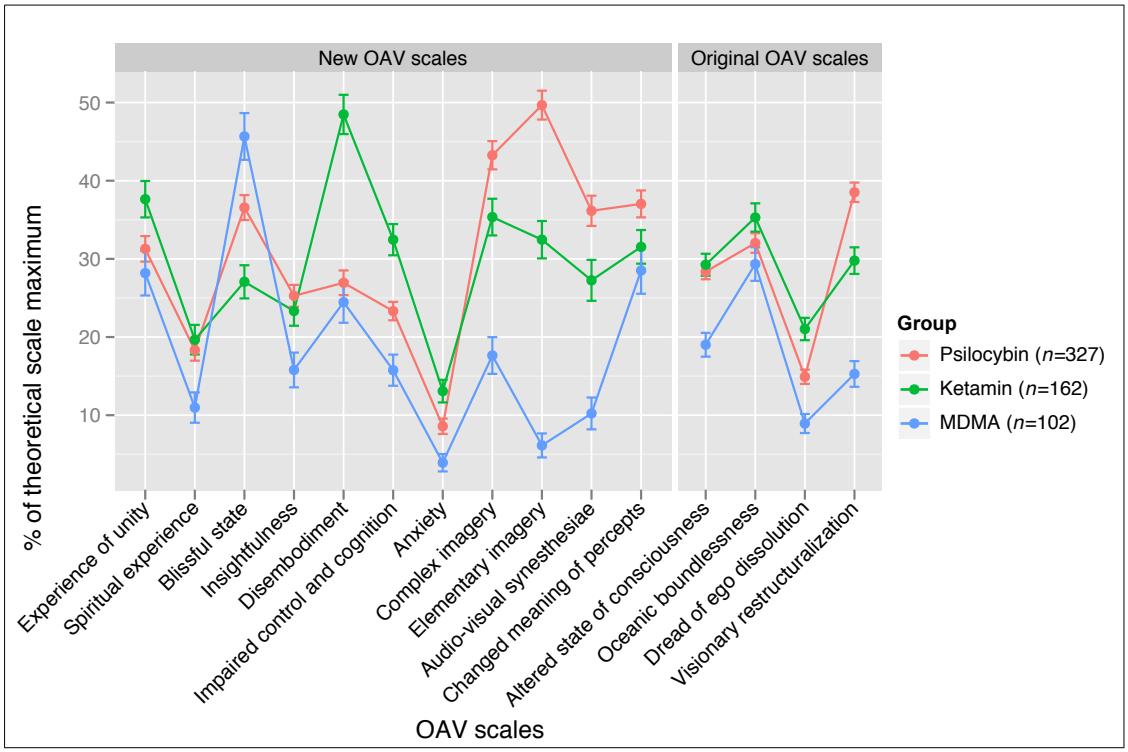

Figure 6: Known-group validities of the original and new OAV scales. Error bars represent standard errors. Double click here to extract the data underlying this plot.

publication. Habermeyer (1999), who conducted a principal component analysis with varimax-rotation on a sample of 93 endogenous psychotic patients who completed the OAV by referring to their most recent acute psychotic episode, found that 10 of the 18 VRS items loaded highest on the OBN factor. Similarly, in a study of Bodmer (1999), in which an EFA with target rotation was conducted using measurements of 135 experimentally induced ASCs, 13 VRS items loaded highest on the OBN factor.

The VRS items that were wrongly assigned to the OBN factor in the studies of Habermeyer (1999) and Bodmer (1999), as well as in the present study, are highly congruent. That is, in all three studies, these items describe experiences of changed meaning of percepts, facilitated recollection, and insightfulness. Additionally, in the studies of Habermeyer (1999) and Bodmer (1999), the wrongly assigned VRS items included items measuring vivid imagery. It should be noted that – except for changed meaning of percepts – these facets were not part of the original conceptualization of the VRS dimension (i.e., in the original APZ questionnaire), but were introduced during construction of the OAV. Because analyses of the APZ had indicated that the VRS dimension describes not only changes in visual perceptions and their associated meanings, but also a general increase in the perception of internally produced stimuli, Bodmer (1989) hypothesized that the VRS dimension could be conceptually extended by incorporating items measuring an increase of imagery thoughts, associations, and memory retrieval. Bodmer's re-conceptualization of the VRS dimension was mainly driven by theoretical considerations of Leuner (1962, 1981), who had speculated that hallucinogenic drugs elicit visual hallucinations by intensifying internal imagery such that the distinction between internally produced imaginary images and external perceptions becomes

blurred. However, given that three studies, including the present study – which has a much larger sample size than the original validation study – have not supported this hypothesis, it appears now that the re-conceptualization of VRS has worsened rather than improved its psychometric properties.

Although reducing the VRS dimension to a set of items tapping only visual alterations would markedly increase its homogeneity, our results indicate that such a construct would still be difficult to separate from the OBN dimension on a high level of the construct hierarchy – especially when potential method effects of similarly worded VRS items are taken into account. Whereas in the three-factorial ESEM without method effects, VRS emerged as a separate factor, it completely merged with the OBN factor when method effects were accounted for by specifying correlated errors. Similarly, the ICLUST algorithm, which seems to be less sensitive to method effects than EFA (Cooksey & Soutar, 2006), combined the OBN and VRS factors to one cluster. This suggests that the VRS factor – at least in part – could be an artifact of method effects. Unfortunately, previous OAV validation studies did not consider this possibility because they exclusively relied on EFA, which cannot account for method effects (Brown, 2006).

Before discussing the issue of what would be the most appropriate number of factors to extract from the OAV, it should be noted that psychological constructs have a hierarchical structure such that different constructs have different levels of conceptual breadth. Hence, the number of factors that can be proposed and assessed is infinite (Reise et al., 2000). The appropriate number of factors to extract depends on the appropriate conceptual breadth of a factor, which, in turn, depends on its specific use. For instance, factors on a high level of the construct hierarchy (i.e., broad constructs) are best suited to predict heterogeneous/complex criteria, whereas narrow-band factors are most efficacious in predicting a specific criterion (Reise et al., 2000). The authors of the OAV decided to extract factors only on a high level of the construct hierarchy because they were primarily interested in the so called etiology-independent dimensions (Dittrich, 1985; Bodmer, 1989). However, even if only higher order factors are considered, we have not found evidence for a parsimonious fit of a three-factorial solution. The ICLUST procedure indicated that only two factors account for the variance between OAV items on a high level of the construct hierarchy. Whereas one of these two factors was equal to the original DED factor, the other consisted of OBN and VRS items. This suggests that, on a high level, the OAV items are best divided on the basis of whether they describe pleasant (OBN and VRS) or unpleasant (DED) experiences. Revelle's VSS criterion, as well as indexes of general factor saturation, such as Revelle's β and McDonald's ω_h, indicated that the OAV items could be combined on an even higher level of the construct hierarchy to form a total scale. This finding is in agreement with the originally proposed general factor G-ASC, which is supposed to be a general measure of the alteration in consciousness (Bodmer, 1989). According to ω_h, the general factor accounted for as much as 65% of the common variance between all 66 items of the total scale. Thus, although the total scale is multidimensional and therefore forms ambiguous correlations with other psychological constructs, the general factor saturation is high enough to justify its use for the prediction of complex criteria (Revelle, 1979; Rossiter, 2002). The same is true for the OBN, DED, and VRS and the "pleasant" and "unpleasant" scales, which also showed strong general factor saturations despite clear rejection of unidimensionality by CFA.

Although the authors of the OAV have considered lower order scales as unreliable and unstable and therefore refrained from their extraction, this study has demonstrated that a number of lower order scales can be constructed that are not only reliable, but also stable (measurement invariant) and valid. Specifically, by using ICLUST, CFA, and MIMIC, we constructed and evaluated 11 new OAV scales formed on the basis of 42 items. The new OAV scales were demonstrated to have many advantages over the old OAV scales. Most importantly, the new scales met criteria of unidimensionality and therefore are more homogeneous than the old scales. Unlike the old scales, the new scales provided a reasonably good fit to the data when modeled as congeneric factors in a simple structure CFA. This is important because a well fitting CFA model is a prerequisite for further analyses within the SEM-framework. For instance, testing measurement invariance by MIMIC or multiple group CFAs and directly estimating reliabilities and disattenuated correlations with other constructs is not possible without a well fitting basic measurement model (Brown, 2006). By using a MIMIC model with five binary predictors, the new OAV factors were demonstrated to be highly measurement invariant across three drug groups, two settings, two questionnaire versions and sexes. Although a small number of items showed DIF, especially when comparing the MDMA and ketamine groups, the impact of DIF on the comparisons of latent factor means was small. This is important for the use of these scales in applied research because it suggests that group mean differences in these scales are hardly confounded by structural differences when calculated on the basis of raw sum scores and by using methods that cannot account for DIF, such as ANOVA.

Although the new scales, due to their lower item number, were less reliable than the old scales when reliability is defined as the proportion of variance that is due to all common factors present in a scale, they still showed relatively high reliabilities. Nine scales had reliabilities beyond 0.8 and two scales had reliabilities between 0.7 and 0.8. Because reliability requirements are weaker when scales are used predominantly to compare groups and not for making decisions about individuals (as it is the case with the new OAV scales), reliability indexes of this size can be considered adequate. In fact, it has been argued that increasing reliabilities much beyond 0.8 in basic research is not worth the effort because measurement error attenuates correlations very little at this level (Nunnally & Bernstein, 1994). Indeed, the lower reliabilities of the new OAV scales did not lead to lower correlations with other psychological constructs, such as the subscales of the EWL-60-S. Even though these correlations were based on raw scores, i.e., were not corrected for measurement error, the new OAV scales tended to correlate more strongly than the old OAV scales. This suggests that the lower reliabilities of new OAV scales were more than compensated by their higher homogeneities.

The new scales were also shown to have good convergent and discriminant validities and to differentiate well among the subjective effects of psilocybin, ketamine, and MDMA. For example, in the MIMIC model, 10 of the 11 new factors were significantly affected by at least one drug contrast variable. The effects of the drug contrast variables supported the known group validities of the new OAV scales because the magnitude and direction of the effects were well in line with what is known about these drugs from the scientific literature. Overall, the new OAV scales differentiated better among the three drug groups than the old scales. For example, the very strong effects of MDMA on blissful state and of ketamine on disembodiment would not have been detected by

using the original OAV scales alone because these experiences would have been mixed up with other experiences measured by the OBN scale.

The interpretation of our results with regard to Dittrich's original hypothesis (i.e., ASCs – independent of their means of induction – can be parsimoniously described by the three oblique primary dimensions OBN, DED, and VRS and the secondary dimension G-ASC) is complicated by the fact that we have analyzed an item set that has been pre-selected to be in accordance with this hypothesis. Unlike the items of the APZ, the items of the OAV were selected and worded to maximally load on one of the three hypothesized primary dimensions (see Bodmer, 1989). Consequently, the factorial structure of the OAV is most likely reflecting this item selection and cannot provide independent evidence for the validity of Dittrich's hypothesis. Unfortunately, as mentioned in the introduction, Dittrich's factorial structure of ASCs may not only be specific to the set of items he selected, it may also be dependent on the data analyzing methods he used. Given these rather severe limitations of Dittrich's original investigations and given that the present study has not confirmed that a three-factorial solution provides a parsimonious fit to the data on a high level of the construct hierarchy, even though the analysis was based on a pre-selected set of items, it seems highly premature to postulate three major dimensions of ASCs, let alone to call them etiology-independent.

4.5.1 *Limitations*

Because the sample of the present study was too small to split it in two halves and to perform exploratory and confirmatory analyses on separate data sets, we have not cross-validated our results. It is therefore possible that we have capitalized on chance at least to some degree. Furthermore, measurement invariance and population heterogeneity of the new OAV scales were only examined by MIMIC modeling and not by multiple-group CFA. This means that we were only able to test the invariance of indicator intercepts and factor means and that all other measurement and structural parameters (i.e., factor loadings, error variances/covariances, factor variances/covariances) were assumed to be equal across the levels of the covariates. Studies that use multiple-groups CFA are clearly needed to further establish measurement invariance of the new OAV scales. The invariance of measurement and structural parameters should also be investigated across additional groups of drugs, dosages, ASC induction methods, settings, and languages.

Although the newest version of Dittrich's ASC rating scales (i.e., the 5D-ASC) contains 94 items, this study has only analyzed the 66 items that it shares with the second newest version (i.e., the OAV). Future studies must clarify whether the common variance between the 28 items that are unique to the 5D-ASC is sufficiently well explained by the two hypothesized factors Vigilance Reduction (VIR) and Auditory Alterations (AUA). Since we have shown that the OBN, DED, and VRS scales can be split into many reliable and valid subscales, it is conceivable that the same could be done with the VIR and AUA scales.

4.5.2 *Conclusions and Recommendations*

The present study confirmed that the general factor (G-ASC) accounts for most of the common variance among OAV items. However, our results only partially supported

the hypothesized structure of group factors. Most importantly, we demonstrated that the OBN, DED, and VRS scales are multidimensional constructs that can be split into many reliable and valid subscales. Although the use of the OBN, DED, and VRS scales – due to their relatively strong general factor saturations – might be justified for predicting complex criteria, we believe that our newly constructed subscales should be preferred for most applications because they are only slightly less reliable but much more homogeneous. Hence, they form less ambiguous correlations with other measures, are easier to interpret, and provide important additional information on more specific experiences of ASCs. We especially caution against the use of the VRS factor in its current form, because a large number of its items repeatedly loaded higher on the OBN than on the VRS factor and because its emergence in EFA might be an artifact of method effects.

5

PREDICTION OF PSILOCYBIN RESPONSE IN HEALTHY VOLUNTEERS

5.1 ABSTRACT

Responses to hallucinogenic drugs, such as psilocybin, are believed to be critically dependent on the user's personality, current mood state, drug pre-experiences, expectancies, and social and environmental variables. However, little is known about the order of importance of these variables and their effect sizes in comparison to drug dose. Hence, this study investigated the effects of 24 predictor variables, including age, sex, education, personality traits, drug pre-experience, mental state before drug intake, experimental setting, and drug dose on the acute response to psilocybin. The analysis was based on the pooled data of 23 controlled experimental studies involving 409 psilocybin administrations to 261 healthy volunteers. Multiple linear mixed effects models were fitted for each of 15 response variables. Although drug dose was clearly the most important predictor for all measured response variables, several non-pharmacological variables significantly contributed to the effects of psilocybin. Specifically, having a high score in the personality trait of Absorption, being in an emotionally excitable and active state immediately before drug intake, and having experienced few psychological problems in past weeks were most strongly associated with pleasant and mystical-type experiences, whereas high Emotional Excitability, low age, and an experimental setting involving positron emission tomography most strongly predicted unpleasant and/or anxious reactions to psilocybin. The results confirm that non-pharmacological variables play an important role in the effects of psilocybin.

5.2 INTRODUCTION

Responses to classical hallucinogens, such as psilocybin, strongly vary between and within subjects, even when the drug dose is kept constant (Nichols, 2004; Studerus, Kometer, et al., 2011). It has therefore long been postulated that a large proportion of inter- and intraindividual differences in reactions to hallucinogens is determined by non-pharmacological variables – also often referred to as set and setting. As originally defined by Leary et al. (1963), set refers to the preparation of the subject, his personality structure, and current mood state, whereas setting refers to the the physical, social, and cultural environment in which the drug is taken. Although set and setting influence the psychological effects of any psychotropic substance, including alcohol and nicotine (e.g. see Janke, 1983), the effects of hallucinogens seem to be particularly strongly determined by these conditions (Eisner, 1997; Nichols, 2004). In fact, they are not only said to be influenced by an individual subject's mental state and surroundings, but to pharmacologically amplify the impact of these non-pharmacological factors on human experience (Grinspoon & Bakalar, 1979; Langlitz, 2010).

Since human hallucinogen research has been dormant for almost three decades and has only come to a revival recently (Vollenweider & Kometer, 2010), most of what we know today about non-pharmacological predictors of hallucinogen response is based on a small number of older studies, many of which do not conform to modern methodological standards. Nevertheless, most of these studies suggest that responses to classical hallucinogens are dependent at least to some degree on the personality structure (e.g., Kornetsky & Humphries, 1957; Rinkel et al., 1961; Fischer, Marks, Hill, & Rockey, 1968; Barr & Langs, 1972; Hemsley & Ward, 1985; Dittrich, 1994; Lienert & Netter, 1996; Bresnick & Levin, 2006). Further influencing factors include the mood state immediately before drug intake (e.g., Metzner et al., 1965; Dittrich, 1994), peer-support (Dunsmore & Kaplan, 1997), estimated emotional support (Leary et al., 1963), expectations of the subjects (e.g., Leary et al., 1963; Metzner et al., 1965; Dittrich, 1994), age (Hyde, 1960; Metzner et al., 1965), body morphology (Metzner et al., 1965), size of the group in which the drug is taken (Leary et al., 1963), and drug pre-experiences (Leary et al., 1963; Metzner et al., 1965).

However, most of these studies have obtained only a limited number of potential predictors at a time. Furthermore, almost all of these studies have relied on simple correlations instead of multiple regression to investigate associations between set and setting variables and drug response. Thus, they did not adjust for potentially confounding variables and also could not reveal the order of importance of different variables. The only exception is a study by Dittrich and his colleagues (Dittrich & Lamparter, 1994; Dittrich, 1994), which has used multiple regression to predict responses to *N*,*N*-dimethyltryptamine (DMT), nitrous oxide, and sensory deprivation from a large number of different set and setting variables. Unfortunately, the sample size of the DMT subgroup was relatively small (n = 45), and the study so far has only been published in book chapters.

Given these methodological problems and given that a growing number of investigators are using hallucinogens for experimental and therapeutic purposes (Vollenweider & Kometer, 2010), new investigations on set and setting are both timely and important. Beyond basic research, such investigations could serve the following purposes. First, they can help to improve the safety of controlled experiments using hallucinogens by providing a basis for deciding which subjects to exclude at screening and how to adjust the environment and procedures for minimizing the risk of adverse reactions. Second, they help to better standardize future experiments. For instance, treatment allocation can be improved by stratifying experimental and control groups on the most important non-pharmacological predictors and efforts in controlling confounding variables can be better directed to those that really matter. Furthermore, the most important predictors can be used for covariate adjustment in randomized controlled trials, which improves precision and power in the estimation of treatment effects (Steyerberg, 2009). Last but not least, knowledge about non-pharmacological predictors can significantly advance our understanding of the neurobiological systems involved in the actions of hallucinogens. This is because individual differences in personality, demographic characteristics, mood, etc. on the one hand, and responsiveness to hallucinogens on the other hand, could be both related to structural and functional differences in specific neurotransmitter systems. In the case of psilocybin, differences are most likely related to differential functioning and density of cortical 5-HT_{2A} receptors because this is the main site of action of

classical hallucinogens (González-Maeso, Weisstaub, et al., 2007; Geyer & Vollenweider, 2008). However, other receptors (particularly the 5-HT_1, 5-HT_4, 5-HT_5, 5-HT_6, and 5-HT_7 receptors) and neurotransmitter systems (particularly the glutamate system), which are additionally involved in the actions of classical hallucinogens (González-Maeso & S. Sealfon, 2009; Vollenweider & Kometer, 2010), might also contribute to common individual differences.

Thus, to further elucidate the dependency of psilocybin response on set and setting, the present study investigates the relative importance of 24 predictor variables, including age, sex, years of education, body mass index, personality traits, drug pre-experience, mental state before drug intake, psychological distress, experimental setting, and drug dose. The analysis is based on the pooled data of 23 controlled experimental studies. Most of these have been published before as single studies. Additionally, data from eight of the 23 pooled studies (i.e., those carried out between 2000 and 2008) were used in a recent pooled analysis on acute, subacute, and long-term subjective effects of psilocybin (Studerus, Kometer, et al., 2011) and data from 20 studies (i.e., all but the three most recent studies) were used in a recent psychometric investigation of the OAV questionnaire (Studerus, Gamma, & Vollenweider, 2010). However, none of these studies have yet reported about the dependency of psilocybin effects on non-pharmacological predictors.

This study improves on previous predictor studies in several ways. First, the sample size (n = 409) is about four times as large as in the largest previous study (Leary et al., 1963). Second, the predictor variables that we used covered a wide range of potentially important domains, and the effects of these predictors were adjusted for the most important confounders. Third, all outcome variables and most of the predictor variables were measured by validated instruments. Fourth, psilocybin was administered under highly standardized research conditions. Finally, by using modern statistical techniques, such as the bootstrap, more reliable estimates of variable importance were obtained.

5.2.1 *Pooled Studies*

The sample used in the present investigation was obtained by pooling raw data from 23 experimental studies (including pilot studies) involving psilocybin administration to healthy volunteers. The studies were conducted at our research facility between 1992 and 2011 as part of a research program in which psilocybin was used as a tool for pharmacological modeling of core symptoms of schizophrenia and for studying cognitive, perceptual, and emotional processes (Vollenweider & Geyer, 2001; Geyer & Vollenweider, 2008). All studies were approved by the Ethics Committee of the University Hospital of Psychiatry, Zurich, and the use of psilocybin was authorized by the Swiss Federal Office of Public Health, Department of Pharmacology and Narcotics, Berne. To minimize potential risks of psilocybin administration, safety guidelines similar to those recommended by Johnson, Richards, and Griffiths (2008) were followed in all studies.

All pooled studies used placebo-controlled within-subject designs. Depending on the study, subjects received placebo and 1-4 different doses of psilocybin in 2-5 experimental sessions, each separated by at least two weeks to avoid carry-over effects. In six of the pooled studies, subjects also received a receptor blocker (i.e., buspirone, ketanserin,

haloperidol, lamotrigine, and risperidone) alone and in combination with psilocybin. In the majority of the studies (n = 16), the order of drug administration was randomized and double-blind, but some of the earlier studies as well as most pilot studies (n = 7) were open-label trials.

For the present analysis, we only used data from experimental sessions in which psilocybin was administered alone and at a dose of at least 115 µg/kg po. Lower psilocybin doses were excluded because they failed to produce subjective drug effects that were statistically different from placebo (Hasler, Grimberg, et al., 2004; Studerus, Kometer, et al., 2011). The pooled sample included 409 psilocybin administrations and 261 subjects. The administered psilocybin dose ranged from 115 to 315 µg/kg po (M ± SD: 214 ± 63 µg/kg; see also Table 13 for the frequencies of different dose categories).

5.2.2 *Subjects*

Participants of all studies were recruited through advertisement from the local universities and hospital staff and carefully screened before admission to the studies. Exclusion criteria were as follows: Personal or family history of schizophrenia, major depression, bipolar, and borderline personality disorder; personal history of alcohol or illicit drug abuse; neurological disorders; and abnormal blood count, electrocardiogram, or blood pressure. Additionally, most studies excluded subjects with an Emotional lability score in the Freiburg Personality Inventory (FPI; Fahrenberg, Hampel, & Selg, 1978) more than two standard deviations above the mean of a normative sample. All subjects gave their written consent after having received detailed information about the aims of the studies, the experimental procedures involved, and the effects and possible risks of psilocybin administration. Subjects were reimbursed for their time and free to withdraw from the study at any time. Descriptive statistics of the included subjects are presented in Table 12.

5.2.3 *Predictor Variables*

Two groups of predictor variables were used: (1) Predictor variables that only varied between subjects, i.e., were constant across different drug sessions of the same individual, and (2) predictor variables that varied between and within subjects. Predictor variables of the first group were measured at screening and included age, gender, body mass index (BMI), years of education, drug use, psychological problems, and stable personality traits, whereas predictor variables of the second group were either measured at the beginning of each drug session shortly before drug administration (e.g., measures of the present mood state) or determined by the design of the experiment (e.g., drug dose, environment of the drug session, and time of assessment). Predictors of the first and second group thus describe subject (Table 12) and session characteristics (Table 13), respectively.

5.2.3.1 *Drug use and pre-experience with classical hallucinogens*

Depending on the study, information on present and past drug use was obtained by semi-structured psychiatric interviews or by one of two different versions of investi-

Table 12: Descriptive Statistics of Subjects (n = 261)

Characteristics	Values[a]	Missings
Age	27.8±6.0	0%
Body mass index	22.2±2.3	25%
Gender		0%
male	62% (161)	
female	38% (100)	
Education		0%
High school diploma	9% (23)	
University students	56% (147)	
University graduates	35% (91)	
Hallucinogen-naïve		15%
yes[b]	59% (131)	
no	41% (90)	
Daily smoker		24%
yes	30% (59)	
no	70% (139)	
THC use		16%
never	16% (35)	
rarely[c]	50% (109)	
sometimes[d]	35% (76)	
Alcohol consumption		23%
<= 60 ml per month	55% (110)	
>60 ml per month	45% (90)	
ZKPQ		
Impulsive Sensation Seeking[e]	0.4±0.8	52%
Neuroticism-Anxiety[e]	-0.9±0.7	52%
Aggression-Hostility[e]	-0.6±0.9	52%
Activity[e]	0.0±0.9	52%
Sociability[e]	-0.1±0.9	52%
TAS		
Absorption[f]	-0.8±1.2	72%
SCL-90-R		
Global Severity Index[g]	-0.3±0.9	31%

Note. THC = Tetrahydrocannabinol; ZKPQ = Zuckerman-Kuhlman Personality Questionnaire; TAS = Tellegen Absorption Scale; SCL-90-R = Symptom Check-List-90-Revised.
[a]Means ± standard deviations and frequencies are shown for continuous and categorical variables, respectively. Numbers in parenthesis indicate absolute frequencies. [b]Experience of a classical hallucinogen at least once in a lifetime previous to the first experimental day. [c]Less than once per month. [d]1-10 times per month. [e]Normed on the Bielefeld-Jena sample (n = 141) of Angleitner, Riemann, and Spinath (2004). [f]Normed on the sample of Ritz and Dahme (1995). [g]Normed on a German community sample (n = 1006; G. H. Franke, 1995).

gator constructed questionnaires. The following categorical predictor variables were constructed by pooling information from all available sources: (1) "Daily-smoker" is a dichotomous variable that is one if the subject smokes at least one cigarette a day and zero otherwise. (2) "Alcohol consumption" is a dichotomous variable that is one if the subject drinks more than 60 ml pure ethanol from alcoholic beverages per month and zero otherwise. (3) "THC use" is an ordered categorical variable with the three categories "never" (absolutely no experience with THC), "rarely" (less than once per month), and "sometimes" (at least once per month). THC use was represented in all regression models as two dummy coded contrast variables using an ordinal coding scheme. That is, when both variables were contained in the model, the first dummy variable represented the difference between "never" and "rarely" and the second between "rarely" and "sometimes". (4) "Hallucinogen-naïve" is a dichotomous variable that is one if the subject has never used classical hallucinogens, such as psilocybin, LSD, and mescaline and zero otherwise. Hallucinogen-naïve is the only drug use variable that could change from one session to another within the same individual because participations on earlier experimental psilocybin sessions were also counted as lifetime hallucinogen experiences. Distributional characteristics of the four drug use variables are displayed in Table 12.

5.2.3.2 *Zuckerman-Kuhlman Personality Questionnaire (ZKPQ; Zuckerman, 2002)*

The ZKPQ contains 99 self-referent true/false statements that cover five major dimensions of personality (1) Impulsive Sensation Seeking consists of of the two facets Impulsivity (i.e., a lack of planning and tendency to act quickly on impulse without thinking) and Sensation Seeking (i.e., a general need for thrills and excitement and preference for unpredictable situations and friends). (2) Neuroticism-Anxiety describes emotional upset, worry, fearfulness, obsessive indecision, lack of self confidence, and sensitivity to criticism. (3) Aggression-Hostility reflects a readiness to express verbal aggression; rude, thoughtless or antisocial behavior; vengefulness; spitefulness; and a quick temper and impatience with others. (4) Activity consists of the two facets Need for General Activity (i.e., impatience and restlessness when there is nothing to do) and Work Activity (i.e., a preference for challenging and hard work). (5) Sociability comprises the two components Parties and Friends (i.e., a liking for big parties, interacting with many people and having many friends) and Isolation Intolerance. There is also a control scale, the Infrequency scale, that serves to eliminate subjects with possibly invalid records.

The ZKPQ is the standard instrument for the assessment of Zuckerman's alternative Five-Factor Model (FFM) of personality. In contrast to the classic FFM (the so called "Big Five"), which has been identified by lexical analyses of words describing personality, the development of the alternative FFM was guided by the assumption that basic personality traits are those with a strong biological-evolutionary basis (Zuckerman, 2002). Consequently, the primary dimensions of the ZKPQ were identified by factor analyzing scores on a variety of personality and temperament scales with known or suspected biological determinants. However, despite conceptual and methodological differences, joint factor analyses of the ZKPQ with the NEO-PI-R (Costa & McCrae, 2008), a well established measure of the Big Five, suggest a large overlap between the two FFMs (P. Schmitz, 2004). That is, the ZKPQ factors Sociability and Neuroticism-Anxiety are considered highly convergent with the Big Five factors Extroversion and Neuroticism, respectively, and

Table 13: Descriptive Statistics of Psilocybin Sessions (n = 409)

Characteristics	Values[a]	Missings
Psilocybin dose (µg/kg)	214.1±63.0	0%
Psilocybin dose (categorized)		0%
115-125 µg/kg	23% (93)	
170 µg/kg	9% (35)	
215-225 µg/kg	20% (83)	
250-270 µg/kg	38% (157)	
315 µg/kg	10% (41)	
Time of assessment[b]		0%
60-90 min	23% (96)	
110-160 min	39% (158)	
195-270 min	23% (94)	
6-10 h	13% (53)	
24 h	2% (8)	
Setting		0%
PET[c]	12% (51)	
no PET	88% (358)	

[a]Means ± standard deviations and frequencies are shown for continuous and categorical variables, respectively. Numbers in parenthesis indicate absolute frequencies. [b]Completion of OAV or 5D-ASC questionnaire after drug intake. [c]Drug sessions involving positron emission tomography measurements.

the ZKPQ factors Impulsive Sensation Seeking and Aggression Hostility have shown at least moderate negative correlations with the Big Five factors Conscientiousness and Agreeableness, respectively.

The authorized German adaptation of the ZKPQ, which has shown good psychometric properties in two independent German samples (Angleitner, Riemann, & Spinath, 2004), was used in 11 of the 23 pooled studies and completed by 125 subjects.

5.2.3.3 *Freiburg Personality Inventory (FPI; half form B; Fahrenberg et al., 1978)*

The FPI half form B contains 105 self-referent true/false statements which – according to the authors of the instrument – measure nine primary and three secondary dimensions of personality. It was administered as part of the screening procedure in 16 of the 23 pooled studies and completed by 155 subjects. The FPI measures very similar personality traits as the ZKPQ. Thus, in order to reduce multicollinearity and to keep the number of candidate predictors low, we only used the scales of the FPI for imputing missing values of the ZKPQ, but not for predicting acute drug responses directly (see statistical analysis section for additional details). Even though the FPI was more completely assessed (40.6 % missings in the FPI vs. 52.1 % missings in the ZKPQ), we decided to use the ZKPQ and not the FPI for predicting psilocybin responses because the ZKPQ has undergone more

extensive psychometric testing and is much more widely used internationally than the FPI half form B.

5.2.3.4 *Tellegen Absorption Scale (TAS; Tellegen & Atkinson, 1974)*

The TAS is a widely used self-report questionnaire for assessing the personality trait absorption. As measured by the TAS, absorption refers to an individual's openness to a variety of cognitive, perceptual, and imagistic experiences as well as vivid imagery, synesthesiae, and intense involvement in aesthetics and nature. The TAS has been reported to be strongly associated with fantasy proneness, and modestly with the Big Five factor Openness to Experience and hypnotic susceptibility (Roche & McConkey, 1990)

We used the German Version of the TAS (Ritz & Dahme, 1995) with a modified item response format (i.e., five-point Likert scale ranging from *does not apply* (0) to *does fully apply* (4) instead of the original dichotomous *true* or *false* response), which is the same version as Ott, Reuter, Hennig, and Vaitl (2005) have used. It was administered in 4 of the 23 pooled studies and completed by 73 subjects. The internal consistency as well as the general factor saturation of the TAS in our sample were excellent (Cronbach's α = 0.95; McDonald's ω_h = 0.75).

5.2.3.5 *Passive-Spontaneous Imagination (PASI)*

The PASI is a subscale of the Hallucination Prediction Inventory (HPI-81; Diezi, Faeh, & Hermann, 1982), which was constructed to explain individual differences in experiencing visual alterations during ASCs. It consists of 30 four-point Likert scale items measuring the frequencies of visual phenomena that spontaneously occur during hypnagogic and hypnopompic states, daydreaming, closed eyes, listening to music, thinking, and imagining (see Supplementary Table S9, for an English translation of the PASI items). The PASI was reported to have good psychometric properties in a normative sample of 442 subjects (Diezi et al., 1982). Furthermore, in a experiment, in which ASCs were induced by sensory deprivation (n = 35), the PASI was a strong predictor of visual hallucinatory phenomena as measured by the Visionary Restructuralization scale of the APZ questionnaire (Dittrich, 1998).

The PASI was administered in 8 of the 23 pooled studies and completed by 107 subjects. The internal consistency of the PASI in our sample was excellent (Cronbach's α = 0.93), and the general factor saturation was satisfactory (McDonald's ω_h = 0.65). There was also a strong correlation of the PASI with the TAS (r = 0.77, n = 53), suggesting a large overlap between these two constructs. In order to reduce redundancy, we only used the PASI to impute missing values of the TAS, but not for predicting psilocybin responses directly. Similar to the FPI and ZKPQ, the TAS was preferred over the PASI even though it had more missing values (72 % missings in the TAS vs. 59 % missings in the PASI) because it is more widely used internationally and has been more extensively validated.

5.2.3.6 *The Symptom Check-List-90-Revised (SCL-90-R; Derogatis, 1994; German version by G. H. Franke, 1995)*

The SCL-90-R is a widely used self-report inventory designed to screen for a broad range of psychological problems present in the past four weeks. Each of the 90 items is rated

on a five-point Likert scale of distress ranging from *not at all* (0) to *extremely* (4). The items of the SCL-90-R are assigned to 9 different symptom dimensions: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Anger-Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. However, because these nine dimensions are not supported by most exploratory and confirmatory factor analyses and because many studies have pointed to the presence of a strong general factor (N. Schmitz et al., 2000), only the Global Severity Index (GSI), which is the total score of all SCL-90-R items, was used as a predictor variable.

The SCL-90-R was administered as part of the screening procedures in 14 of the 23 pooled studies and completed by 179 subjects. The internal consistency of the GSI in the pooled sample was excellent (Cronbach's $\alpha = 0.94$), and the general factor saturation was satisfactory (McDonald's $\omega_h = 0.62$).

5.2.3.7 *Adjective Word Lists ("Eigenschaftswörterliste"; EWL-60-S and EWL-K)*

Two different versions of the Adjective Word List were used to assess the current mental state shortly before drug administration. The older version EWL-K (Janke & Debus, 1978) was used in studies before the year 2000 ($n = 7$), whereas the newer version EWL-60-S (Janke & Debus, 1986) was used in later studies ($n = 5$). Both questionnaires contain a list of adjectives which must be rated on how well they describe the current mental state. The EWL-K contains 123 adjectives and a dichotomous *true* or *false* response format, whereas the EWL-60-S contains 60 adjectives and a four-point response format ranging from *not at all* (0) to *strongly* (3). In both questionnaires, items are grouped into six main scales: Performance-Related Activity, General Inactivation, Extroversion-Introversion, General Well-Being, Emotional Excitability, and Anxiety-Depressiveness. We combined these scales across questionnaire versions by using only those items that are contained in both questionnaire versions (see Supplementary Table S10 for a list of the overlapping items in each scale). To adjust for the different item response format, each EWL scale was z-transformed within each questionnaire version. By combining EWL-K and EWL-60-S, measures of the mental state before drug intake were available from 185 of the 409 drug sessions.

5.2.4 *Response Variables*

5.2.4.1 *Altered States of Consciousness Rating Scales OAV and 5D-ASC)*

In each experimental session, subjects were asked to rate drug induced alterations of consciousness by either the OAV questionnaire (Bodmer et al., 1994) or its extended version 5D-ASC (Dittrich et al., 2006). The OAV was used in studies conducted before the year 2000 ($n = 10$), whereas the 5D-ASC was used in all later studies ($n = 13$). In all studies, questionnaires were administered during the acute or post-acute effects of the drug and subjects were asked to rate their experiences from the moment of drug intake to the time of assessment. If the questionnaires were completed more than once during an experimental session, only data from the measuring time points yielding the highest mean total score were used. The frequencies of different assessment times in the pooled sample are shown in Table 13. Because the time of assessment could have

affected subjective drug effects ratings, time (defined as the logarithm of minutes after drug intake) was included as covariate in all statistical analyses.

There are 66 visual analogue items that occur in both the OAV and 5D-ASC and that can be used to assess three primary and one global dimension of ASCs. The three primary dimensions are called Oceanic Boundlessness (OBN), Dread of Ego Dissolution (DED), Visionary Restructuralization (VRS), and the global dimension is called Global Altered States of Consciouness (G-ASC). The OBN dimension describes highly enjoyable and positively valued experiences of ASCs, such as deeply felt positive mood, experiences of unity, transcendence of time and space, spiritual experiences, and sense of intuitive understanding. Because many of the OBN items have been directly formulated on the basis of six of the nine categories of mystical experiences proposed by Stace (1961), high scores on the OBN scale indicate a state similar to mystical experiences as described in the scientific literature on the psychology of religion. The DED dimension measures experiences of cognitive impairment, loss of self-control, feelings of disintegration or separation from oneself and the world, and anxiety or panic. High scores on the DED scale therefore indicate a very unpleasant state similar to so called "bad trips" described by drug users. The VRS dimension assesses elementary and complex visual pseudo-hallucinations, audio-visual synesthesiae, increased production of vivid imagery from memory or fantasy, as well as changes in the meaning of percepts. Finally, the secondary scale G-ASC is the total score of all 66 OAV items and thus can be considered as a general measure of consciousness alteration.

The OBN, DED, VRS, and G-ASC dimensions have been hypothesized to be fundamental dimensions of ASCs that are factorially invariant across ASC induction methods (Dittrich, 1998). However, a recent psychometric investigation of the OAV (Studerus, Gamma, & Vollenweider, 2010) has only partially confirmed this hypothesis. Specifically, it has been found that the VRS factor contains several items that load more strongly on the OBN factor and that the VRS factor could be merged with the OBN factor on a high level of construct hierarchy. Furthermore, all original OAV factors were demonstrated to be multidimensional. Studerus, Gamma, and Vollenweider (2010) therefore constructed and validated eleven new lower order factors that are more homogeneous than the original factors and that can be used to describe more specific aspects of ASCs (see Studerus, Gamma, & Vollenweider, 2010, for descriptions of these scales). Nevertheless, because the original factors have shown relatively strong general factor saturations, they still can be advantageous for capturing complex criteria. To this end and in order to compare our results with earlier studies, we decided to use both the original and the recently constructed subscales as dependent variables.

5.2.5 *Statistical Analysis*

To ensure the validity of the assumptions of linear mixed effects models (i.e., Gaussian distribution of random effects and within-subjects errors, homoscedasticity, and linearity), an extended method of Box-Cox transformation (Gurka, Edwards, Muller, & Kupper, 2006) was applied to all response variables. The negative inverse of the square root was found to be appropriate for the Anxiety factor, and a natural logarithm transformation worked best for the DED, Spiritual Experience, Insightfulness, Disembodiment, and Audio-Visual-Synesthesiae factors. All other response variables were

transformed by taking the square root. Predictor variables were not transformed because partial residual plots indicated that linearity assumptions were already reasonably well satisfied after transforming the response variables.

As depicted in Supplementary Figure S3, several predictor variables contained considerable proportions of missing values. Because the missing data mostly resulted from different study designs among the pooled studies, the missing data mechanism can be assumed to be "missing at random" MAR or "missing completely at random" (MCAR; Enders, 2010). To minimize potential bias and loss of information arising form missing data, we used a statistical technique called multiple imputation (MI; Rubin, 1987). MI is regarded as the method of choice for handling complex incomplete data problems because it yields unbiased parameter estimates and standard errors under an MAR or MCAR missing data mechanism and maximizes statistical power by using all available information (Enders, 2010). We imputed missing values by the Multivariate Imputation by Chained Equations (MICE) software (van Buuren & Groothuis-Oudshoorn, 2011), which is freely available as an add-on package to R (R Development Core Team, 2011). The MICE-package uses fully conditional specification as imputation method, which means that imputation models can be flexibly specified on a variable-by-variable basis. We used predictive-mean-matching, proportional odds models, and logistic regressions to impute continuous, ordered categorical, and binary variables, respectively. The scales of the the FPI and PASI questionnaires were included in the MI procedure as auxiliary variables to improve the imputation of the ZKPQ and TAS scales, respectively. For each variable, the set of predictors was restricted to those that correlated with at least 0.15 with the variable to be imputed. This resulted in a series of imputation models that contained the best 9-29 predictors of each target variable. Due to the relatively large fraction of missing information in some variables, we generated 20 multiply imputed data sets, which is a larger number than what the literature historically recommends (Enders, 2010). Recent simulation studies (e.g., Graham, Olchowski, & Gilreath, 2007) show that this has a very beneficial impact on statistical power, especially when the fraction of missing information is as high as in the present study. Convergence of the Gibbs sampling algorithm and the quality of imputed values were assessed in accordance with recommendations of van Buuren and Groothuis-Oudshoorn (2011).

To ensure that no severe multicollinearity existed between predictor variables, variance inflation factors (VIFs) were computed for each predictor variable within in each of the imputed data sets. Because no VIF was larger than 3, we did not exclude any predictor variable due to multicollinearity.

Because some subjects participated in more than one psilocybin study and because some studies involved multiple psilocybin sessions, our pooled data set contains non-independent observations. To account for this non-independency, we used linear mixed models in which the intercepts were allowed to vary per subject. We also considered more complex mixed effects models with varying slopes for the drug dose effects and varying intercepts by study. However, model comparisons by Akaike's information criterion (AIC) in the full models suggested that the varying intercept per subject model was sufficient to account for the clustering in our data.

In order to directly compare regression coefficients of binary and continuous predictors, continuous predictor variables were rescaled within each imputed data set by dividing them by two times their standard deviations. Because binary variables – except

when highly skewed – have a standard deviation of roughly 0.5, our rescaling procedure resulted in regression coefficients that reflected the change of the dependent variable for a two standard deviation change in both binary and continuous predictors (see also Gelman, 2008). Outcome variables were z-transformed within each imputed data set such that regression coefficients were also comparable across models with different outcomes.

Regression models that contain too many unimportant predictor variables can result in loss of precision in the estimation of regression coefficients and the predictions of new responses (Royston & Sauerbrei, 2008). On the other hand, selecting variables by data-dependent methods (e.g., stepwise approaches) may result in overly optimistic estimates of predictive ability and model fit and unstable sets of predictor variables, especially in small data sets (Steyerberg, 2009). To reduce these risks, we built our models by combining backward elimination with a two step bootstrap approach. A major advantage of this approach is that it also solves the problem of variable selection under multiple imputation (e.g., Heymans, van Buuren, Knol, van Mechelen, & de Vet, 2007).

In the first step, 200 bootstrap samples were taken from each of the 20 imputed data sets. The bootstrap samples were obtained by drawing from individual cases (i.e., psilocybin sessions) with replacement and were of equal sample size as the original sample. Within each bootstrap sample, parsimonious models were searched for by applying backward elimination. That is, starting from the full models, predictors were dropped in a stepwise fashion until no predictor was left with a Wald test p-value larger than 0.157. This significance level corresponds to selecting predictors with 1 df based on the AIC (Steyerberg, 2009). The random intercept for the subjects was always included in the models and only fixed effects were considered for elimination. Predictors were ranked according to their inclusion frequencies in the final models. Those predictors that were selected in at least 50% of the 20×200 bootstrap samples were considered important and further analyzed in a second modeling step. Thus, the first step primarily served to reduce model space with a minimal risk of eliminating important predictors (see also Sauerbrei, Holländer, & Buchholz, 2008).

The second modeling step was very similar to the first step. Again, 200 bootstrap samples were taken within each imputed data set and backward elimination was applied with a stopping rule of $p = 0.157$. However, this time we began from full models that included only those predictors that were selected in the first step. The model that was selected most often across all bootstrap samples was considered the most stable model and further explored for relevant interactions. The two step bootstrap procedure was repeated for each of the 15 response variables.

In each of the resulting 15 final models, the multivariate associations between the repeatedly measured outcomes and the fixed effects in the models were estimated by the R^2 statistic proposed by Edwards, Muller, Wolfinger, Qaqish, and Schabenberger (2008). To obtain reliable standard errors, confidence intervals, and associated p-values of the fixed effects parameters, Markov Chain Monte Carlo (MCMC) sampling was used. MCMC sampling is a modern alternative to the conventional significance test of fixed effects in mixed effects models based on t or F statistics, which is unreliable due to the lack of a clear definition of the degrees of freedom (Baayen, Davidson, & Bates, 2008). For each final model and each imputed data set, random draws from the posterior

distributions of the parameters were taken and then mixed across data sets. The mixed draws approximate the posterior distribution of the pooled parameters and thus can be used for inference after multiple imputation (Gelman, 2004). A simulation study by Zhou and Reiter (2010) has shown that this approach leads to better results than the conventional application of Rubins's rules (Rubin, 1987), especially when the number of imputed data sets is large.

5.3 RESULTS

The selection frequencies of the predictor variables and models as a result of the bootstrap selection procedures are provided in Supplementary Table S11. Drug dose was the only predictor variable that reached a selection frequency of 100% and it did so with all 15 response variables. The number of predictor variables that were selected in more than 50% of the bootstrap samples ranged from 5 for the dependent variable DED to 20 for the dependent variable Audio-Visual Synesthesiae. BMI and daily smoker were the only variables that never reached a selection frequency of 50%. In general, the number of predictor variables reaching the cutoff of 50% was considerably lower for the original OAV scales than for OAV subscales (6-7 vs. 16-24). The lower number of selected variables in the first step also led to more stable models in the second step. Whereas the most frequently selected model of the dependent variable OBN was selected in 32.1% of the cases, the most frequently selected model of the dependent variable Anxiety was selected only in 0.3% of the cases.

Because drug dose was clearly the most important predictor, first order interactions between drug dose and all other predictors were explored within the most frequently selected models. Only 12 of the 15 $\times$ 23 tested interactions were significant at $p < 0.05$ and only one interaction, namely, PET $\times$ Drug Dose predicting Anxiety, reached significance at $p < 0.01$. The interaction indicated that Anxiety increased with increasing drug dose in the non-PET condition, but decreased with increasing drug dose in the PET condition. However, it should be noted that the variability of the drug dose variable within the PET condition was very small. Specifically, only 215 and 250 µg/kg doses of psilocybin were administered in experiments involving PET measurements. Thus, the significance of this interaction was considered rather questionable and not included in the final models.

The variances explained in the full and simplified final models of all 15 outcome variables are presented in Table 14. As can be seen from the table, the variable selection procedure only slightly reduced explained variances, suggesting that the most important predictors were retained in the models, and the excluded variables were mostly noise variables. The variance explained in the simplified models was highest for the OAV total scale ($R^2 = 0.31$) and lowest for Disembodiment ($R^2 = 0.163$). In general, the main scales tended to have higher explained variances than the subscales even though they were explained by a lower number of predictors.

The size and statistical significance of the regression coefficients of the most stable models, as estimated by the MCMC sampling method, are shown in Figure 7. Standard errors and highest posterior density 95% credibility intervals are additionally provided in Supplementary Table S11. Overall, drug dose had the strongest effect on psilocybin response. It was significantly associated with all outcome variables and had the highest

Table 14: Variance Explained in the Full and Simplified Models

Outcome	Edward's R^2	
	Full models	Simplified models
Main scales		
Altered State of Consciousness	0.337	0.310
Oceanic Boundlessness	0.293	0.266
Dread of Ego Dissolution	0.232	0.182
Visionary Restructuralization	0.333	0.293
Subscales		
Experience of Unity	0.221	0.196
Spiritual Experience	0.249	0.220
Blissful State	0.195	0.179
Insightfulness	0.225	0.184
Disembodiment	0.175	0.163
Impaired Control and Cognition	0.202	0.192
Anxiety	0.201	0.187
Complex Imagery	0.237	0.229
Elementary Imagery	0.205	0.194
Audio-Visual Synesthesiae	0.226	0.217
Changed Meaning of Percepts	0.234	0.221

effect size in all models, except in the models predicting Spiritual Experience, Anxiety, and Changed Meaning of Percepts. The time of assessment was positively associated with Spiritual Experience and Elementary Imagery and negatively associated with DED and Impaired Control and Cognition, indicating that disturbances in control and cognition were less often reported when asked about later in the session. In experimental sessions involving PET measurements, participants reported much higher levels of Anxiety. In fact, of all 24 analyzed predictor variables, PET was the strongest predictor of Anxiety, and its effect size was more than twice as high as the one of drug dose. Compared to younger subjects, older subjects reported less Impaired Control and Cognition and also showed a trend for more Blissful State ($p = 0.059$). Years of education, gender, and BMI were not significantly associated with any response variable.

Drug use and pre-experiences with hallucinogenic drugs only moderately affected psilocybin responses. Although hallucinogen-naïve subjects tended to report stronger effects in most outcome variables, statistical significance was only reached for Disembodiment, VRS, and Changed Meaning of Percepts. Psilocybin responses did not differ between subjects who never consumed THC and those who rarely consumed THC. However, subjects who sometimes smoked cannabis (more than once per month) reported significantly more Blissful State than subjects who rarely consumed cannabis (less than once per month) and also showed a trend for less Anxiety ($p = 0.07$). There

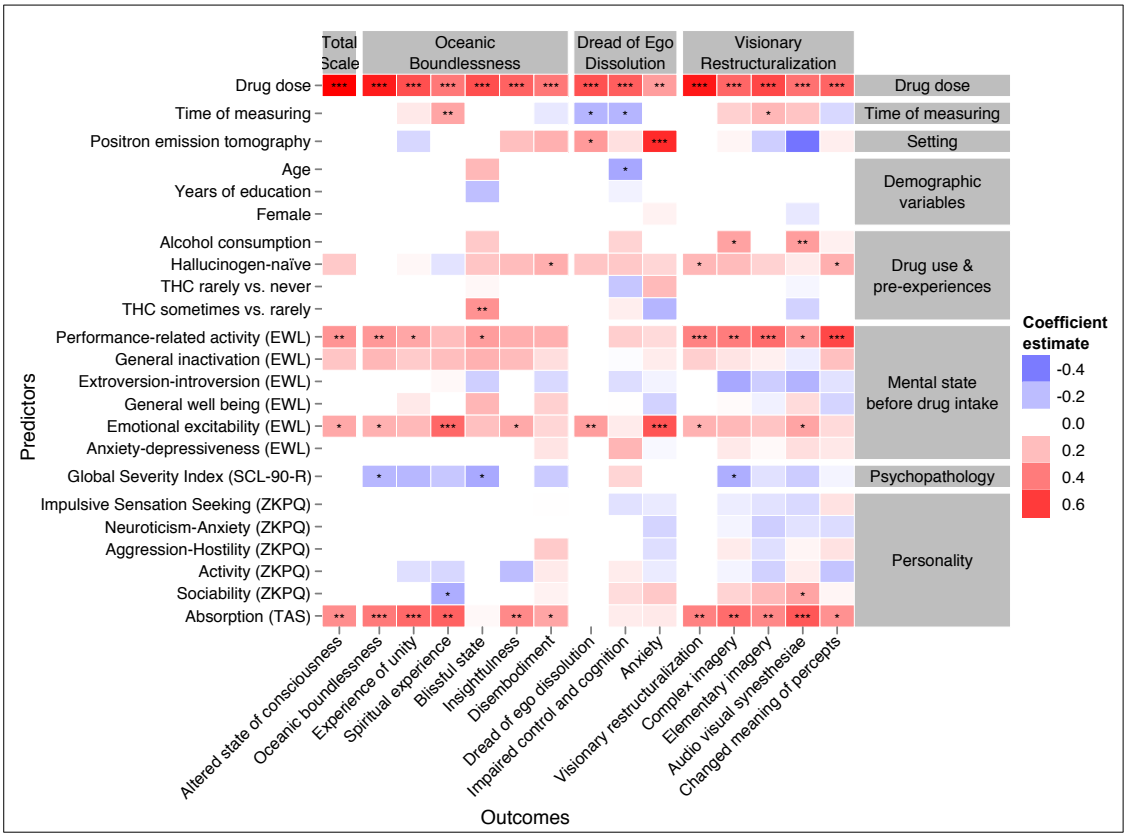

Figure 7: Regression coefficients of the final models pooled across 20 imputed data sets. The effects are adjusted for the influences of all other variables in the models. One, two, and three asterisks represent p-values < 0.05, 0.01, and 0.001, respectively.

were also statistically significant positive associations between alcohol consumption and experience of Audio-Visual Synesthesiae and Complex Imagery.

The mental state immediately before drug intake had a relatively strong influence on several outcome variables. Specifically, Performance-Related Activity, which was measured by the adjectives go-getting, avid, active, and energetic, had a major influence on the overall consciousness alteration (G-ASC) and on several experiences covered by the OBN and VRS dimensions. Emotional Excitability was strongly positively associated with Spiritual Experience and Anxiety and moderately with all OAV main scales, as well as Insightfulness and Audio-Visual Synesthesiae. High Anxiety-Depressiveness before drug intake did not lead to significantly more unpleasant experiences during the sessions.

The GSI scale of the SCL-90-R was negatively associated with OBN, Blissful state, and Complex Imagery, indicating that subjects who experienced more psychological problems previous to the experiments reported less pronounced effects with these scales. Except for Absorption, which strongly predicted several experiences measured by the OBN and VRS dimensions, most personality traits did not have a major influence on psilocybin responses. Of the personality traits constituting Zuckerman's alternative five-factor model, only Sociability was significantly associated with any outcome variable. Specifically, subjects who were more sociable (i.e., outgoing and extroverted) reported less Spiritual Experience and more Audio-Visual Synesthesiae.

The fractions of missing information (FMI) and the relative increases in variance due to missingness (RIV), which quantify the missing data's influence on the sampling variance of the parameter estimates, are shown in Supplementary Table S11. For all imputed variables, FMI values were lower than their missing data rates, which indicates that the variables in the imputation model were predictive of the missing values. Because some of the information loss was mitigated by borrowing information from correlated variables, increases in sampling errors of the regression coefficients were not completely commensurate with overall reductions in sample sizes. Not surprisingly, the RIV was largest for Absorption (0.94 on average), which was also the variable with the highest missing data rate. This indicates that the confidence interval of the regression coefficient for Absorption was on average about 0.94 times larger than it would have been if this scale had no missing values.

5.4 DISCUSSION

The present study sought to predict acute responses to psilocybin when administered in a controlled scientific setting to healthy volunteers. The relative importance of 24 predictor variables from a wide range of domains were investigated.

Drug dose was clearly the most important predictor of psilocybin response. It was the only predictor that was always retained in automatic variable selection and its effect size was largest in 12 of the 15 final prediction models. Furthermore, its effect on general consciousness alteration, as measured by the OAV total scale, was more than twice as high as that of other predictors. The personality trait of Absorption was found to be the second most important predictor of psilocybin response. It was highly positively associated with the overall consciousness alteration and strongly predicted mystical-type experiences and visual effects induced by psilocybin. Further variables that were found to be important for predicting psilocybin response were Performance-Related Activity, Emotional Excitability, psychological distress as measured by the GSI, pre-experience with classical hallucinogens, frequencies of THC and alcohol consumption, Sociability, time of assessment, and setting (PET vs. no PET measurement). Being in an emotionally excitable and active state immediately before drug intake, having experienced few psychological problems in the past weeks, no previous experience with classical hallucinogens, and moderate THC and alcohol consumption increased the intensity of pleasurable effects and/or visual alterations, whereas settings involving PET measurements, Emotional Excitability, and low age contributed to the experience of unpleasant and/or anxious reactions.

The finding that Absorption was amongst the most important predictors of psilocybin-induced ASCs is consistent with a large number of studies showing that Absorption is associated with differential responsivity to various ASC induction procedures, including hypnosis, meditation, marijuana intoxication, and electromyograph biofeedback (Pekala, Wenger, & Levine, 1985; Vaitl et al., 2005). Absorption has also been reported to be positively associated with the occurrence of synesthesiae after the ingestion of ayahuasca (Bresnick & Levin, 2006), a hallucinogen with similar modes of action as psilocybin. This is in agreement with our results, which showed that, of all 15 response variables, Absorption most strongly predicted Audio-Visual Synesthesiae. A recent study by Ott et al. (2005) suggests that inter-individual differences in Absorption and

responsiveness to hallucinogenic drugs could be both related to the binding potential of the 5-HT_{2A} receptor, which is the main site of action of serotonergic hallucinogens, such as psilocybin (Vollenweider & Kometer, 2010). Although Ott et al. (2005) have demonstrated a significant association between T102C polymorphism affecting the binding potential of the 5-HT_{2A} receptor and the TAS scale, they did not assess the association between TAS and responsivity to serotonergic hallucinogens. The present study is filling this gap, as it is, to our knowledge, the first study predicting the effects of a classical hallucinogen by Absorption in a larger sample.

Apart from a strong influence of Absorption and a relatively minor influence of the ZKPQ factor Sociability, which is highly convergent with the Big-Five factor Extroversion (P. Schmitz, 2004), personality traits only marginally contributed to the prediction of psilocybin responses. This is rather surprising because personality traits have been postulated by many authors to be among the most important determinants of hallucinogen response (e.g., Fischer, Marks, et al., 1968; Barr & Langs, 1972). It is also worth noting that we did not detect any statistically significant relationship between Neuroticism-Anxiety and negative reactions to psilocybin. This finding contradicts several earlier, smaller scaled studies (Hemsley & Ward, 1985; Dittrich, 1994; Lienert & Netter, 1996), which have found moderate to strong correlations between Neuroticism and anxious reactions to classical hallucinogens and which have led to our policy of excluding subjects with very high Neuroticism scores (i.e., more than two SD above the mean) at screening. Although the exclusion of highly neurotic subjects could have distorted our sample and thus reduced the predictive ability of Neuroticism, it should be noted that the chosen cutoff affects only the highest 2.3% of the normal distribution and that there was still substantial variability of Neuroticism in our sample. Nevertheless, as has been shown in Table 12, both mean and variance of Neuroticism in our sample where somewhat reduced compared to normative data. Hence, we cannot rule out the possibility that Neuroticism increases the risk of adverse reactions in the highest tail of the distribution. A positive relationship between Neuroticism and the effects of classical hallucinogens would also be biologically plausible because a recent PET study has demonstrated a positive correlation between Neuroticism and frontolimbic 5-HT_{2A} receptor binding (Frokjaer et al., 2008). Thus, excluding highly neurotic subjects at screening might still be a sensible approach for increasing the safety of controlled experiments involving hallucinogen administration.

In contrast to personality factors, current mood state and psychological distress in the past four weeks before drug intake were generally more important for predicting psilocybin response in this study. This is in agreement with the existing literature. For instance, Metzner et al. (1965) have found that the best predictor for mood during the psilocybin session was mood before the session, and Dittrich (1994) reported that Emotional Lability, a factor that was predominantly measured by state variables, most strongly increased the likelihood of experiencing DED after DMT administration. Interestingly, we have found that Emotional Excitability shortly before drug intake predicted anxious reactions to psilocybin much better than Anxiety-Depressiveness. However, this could also be due to statistical reasons. Whereas Emotional Excitability was measured by 11 items, Anxiety-Depressiveness was measured by only 4 items that additionally also had relatively high item difficulties. Consequently, the variability – and possibly also the reliability – of Anxiety-Depressiveness was substantially lower

than that of Emotional excitability. It should also be noted that these two factors were relatively highly correlated in our sample (r = 0.5), which frequently might have led to the inclusion of only one of these two variables in automatic model selection. The finding that Performance-Related Activity was amongst the most important predictors of experiences described by the OBN and VRS dimensions has, to our knowledge, not been described in the literature before. One possible explanation is that the items assessing Performance-Related Activity (i.e., go-getting, avid, active, and energetic) not only captured variance associated with fitness and energy, but also variance with positive mood and general optimism. Correlations with EWL subscales (not reported) are in support of this hypothesis because they reveal that Performance-related activity is most strongly associated with the EWL subscale Heightened mood (r = 0.47).

The finding that the PET environment was strongly associated with anxious reactions could be partially explained by the perceived atmosphere at the PET center. Whereas non-PET experiments were mostly conducted in laboratory rooms that were furnished in an aesthetically pleasing way, the environment at the PET center was much more clinical and "antiseptic" (i.e., lots of technical equipment, white walls, personnel in white lab coats). Our results are therefore in support of current safety guidelines (Johnson, Richards, & Griffiths, 2008), which recommend avoiding "cold" and overly clinical environments in human hallucinogen research in order to reduce the risk of anxious reactions. Although we have found increased Anxiety in PET experiments, that does not mean that psilocybin experiments involving PET measurements are unsafe. The percentage of strong anxious reactions in the PET experiments was still relatively low, and all of them could be successfully managed by providing interpersonal support. Furthermore, there are other factors that might have contributed to the increased Anxiety in the PET environment. For instance, in contrast to non-PET experiments, subjects could have their eyes closed while lying in the scanner and they were less distracted by performing tasks. Thus, they could concentrate more on the experience, which in turn might have increased the confrontation with inner fears.

Our results indicate that, in contrast to MDMA (Liechti, Gamma, & Vollenweider, 2001), the effects of psilocybin were not moderated by gender. This is consistent not only with earlier studies investigating the subjective effects of classical hallucinogens in humans (Hyde, 1960; Leary et al., 1963), but also with neuroimaging studies, which have found no gender differences in 5-HT_{2A} receptor binding in cortical regions (Adams et al., 2004; Frokjaer et al., 2008). The only demographic variable that was statistically significantly associated with any psilocybin response in this study was age. Specifically, older subjects reported less Impaired Control and Cognition and tended to experience more Blissful state compared to younger subjects. These associations are similar to those observed by Hyde (1960) and Metzner et al. (1965) and could be explained by an increased experience with managing occurrent negative emotions in older people (Blanchard-Fields, 2007). It is also consistent with the fact that 5-HT_{2A} receptors densities decrease with increasing age (e.g., Adams et al., 2004).

The finding that hallucinogen-naïve subjects reported slightly more VRS, Disembodiment, and Changed Meaning of Percepts is consistent with the study of Metzner et al. (1965), which found that previous experience with classical hallucinogens was negatively associated with the number of somatic symptoms and visual alterations induced by psilocybin. However, our results disagree with those of Dittrich (1994),

who found that familiarity with drug-induced ASCs was not predictive of any acute effects of DMT, as measured by the OAV questionnaire. One possible explanation of this discrepancy might be that the predictor variable "familiarity with drug-induced ASCs" in the study of Dittrich not only included pre-experience with classical hallucinogens, but also other psychotropic substances.

Although subjects were asked to rate their experiences in retrospect and mostly during or after the peak effects of the drug, Impaired Control and Cognition induced by psilocybin was rated as less intense and Spiritual Experience and Elementary Imagery were rated as more intense when questionnaires were completed later in the sessions compared to earlier in the sessions. These findings are in agreement with a study of Linton, Langs, and Paul (1964), which found that subjects tended to forget ego-alien and threatening aspects of an LSD experience more often than those dealing with affects or changes in the perceived meaning of events. Although it is tempting to explain these associations by the well-known phenomenon of "motivated forgetting" (Anderson & Levy, 2009), they could also have resulted from differential time courses of psilocybin effects (e.g., see Studerus, Kometer, et al., 2011).

5.4.1 *Limitations*

The present study has several limitations. First of all, the design of our study does not allow causal interpretations of predictor effects. Although we analyzed data from experimental studies, the only variable that was systematically manipulated was drug dose and only within, not between, studies. Hence, with the exception of drug dose, associations between predictors and outcomes are purely observational.

Although several statistically significant relationships between non-pharmacological predictors and outcome variables were detected, there were still relatively large proportions of unexplained variances in the outcome variables. For instance, more than 80% of the variance of the outcome variable Anxiety was left unexplained, suggesting that there is considerable unpredictability in anxious reactions to psilocybin – even under highly standardized conditions.

Generalizations of our results are hindered by the composition of our sample and the circumstances in which psilocybin was administered. For instance, our subjects were relatively young, highly educated, and high-functioning. They had more pre-experiences with classical hallucinogens and cannabis than their corresponding age group in the general population and also showed low Neuroticism-Anxiety scores (i.e., almost one SD below the mean of a normative sample). The distortions in our sample most likely resulted from our recruitment method, which is prone to self-selection bias (see also Studerus, Kometer, et al., 2011). The specific composition of our sample and the fact that psilocybin was administered in a carefully monitored research environment might have reduced the occurrence of unpleasant reactions (i.e., so called bad or horror trips). This in turn might have lowered our ability to detect risk factors for unpleasant reactions.

The individual studies that were pooled for the present analysis were not specifically designed to investigate predictors of psilocybin response. Consequently, the predictor variables analyzed herein are not necessarily those that – according to the literature – would be most promising to investigate. Although the studied predictors cover the

most important domains, some of them are clearly underrepresented. For instance, the influence of the setting was only covered by the PET vs. no PET variable. Furthermore, expectancies of the subjects, which are well known to influence the effects of most psychoactive drugs, including alcohol and nicotine (Vogel-Sprott & Fillmore, 1999), could not be studied because no such variables were obtained.

Because the majority of the pooled studies used double-blind placebo-controlled designs, one might argue that we could have controlled for expectancy effects by including the response to placebo as a covariate into the analyses. Unfortunately, the effects of psilocybin were so strong that most subjects could easily differentiate them from placebo. Moreover, because the items of the OAV questionnaire are visual analogue, anchored *no, not more than usual* on the left and *yes, much more than usual* on the right, most subjects placed marks at the left end of the scale for all items once they were convinced that they had received placebo. Consequently, mean and variances of the OAV scales were essentially zero under placebo, which severely limited the usefulness of these scales as covariates. While some investigators have used an active placebo to increase the success of the double blind in experiments involving hallucinogens (e.g., Griffiths, Richards, McCann, & Jesse, 2006), an even better approach for separating pharmacological effects from the cognitive expectations of receiving the drug and its effect might be the so called balanced-placebo design (BPD; Marlatt & Rohsenow, 1980). The BPD is a 2×2 factorial design that crosses the administered substance (drug vs. placebo) with an instructional set manipulation (subjects are told they receive the active drug vs. subjects are told they receive placebo). To our knowledge, the BPD has not yet been used in experiments with classical hallucinogens, but a recent study has demonstrated its feasibility with marijuana (Metrik et al., 2009). It is therefore conceivable that the BPD could also foster our understanding of expectancy effects in responses to classical hallucinogens.

Another limitation of the present investigation is that responses to psilocybin, as measured by the OAV, could be confounded by individual differences in the interpretation of the item anchors at the right end of the visual analogue scale. Specifically, the anchor *yes, much more than usual* could have had different meanings depending on whether the subject has experienced profound ASCs before. Future studies should therefore validate our results by also using behavioral measures and/or external raters for assessing psilocybin response.

In the present study, we have only predicted single aspects of ASCs. Another approach, taken by Barr and Langs (1972), is to predict patterns of psilocybin responses. This could be accomplished by cluster analyzing individual responses using Pearson correlations as a proximity measure. The response clusters could than be predicted by multinomial regression models. Because psilocybin, especially with higher doses, sometimes can elicit responses that are not only quantitatively but also qualitatively different (Nichols, 2004), it is possible that a categorical approach would be better suited to detect determinants of profound ASCs, such as mystical-type experiences or so called "horror trips". The main reason why we did not follow such an approach in this investigation is that these experiences only occurred in a small proportion of our subjects (cf. Studerus, Kometer, et al., 2011). Hence, even with our large sample, the event-per-variable ratio and statistical power were considered too low for such an analysis.

A few further statistical issues are worth noting. Although we used a two-step bootstrap procedure to protect against the dangers of data-driven model selection, the stability of some prediction models were relatively low. For instance, the most frequently selected model of the outcome variable Anxiety was selected in only 0.3% of the cases, and there were many other models that were only slightly less frequently selected. Thus, there was considerable uncertainty in some of the final models, which could have introduced bias in the estimation of regression coefficients and confidence intervals (Steyerberg, 2009). The natural remedy for this problem would have been to base inference on a set of competing models using model averaging and the selection frequencies as model weights (Sauerbrei et al., 2008). However, because our analysis was already complicated by the fact that we had used mixed effects models in combination with multiple imputation, we did not want to introduce additional complexity into the analysis and therefore abstained from performing frequentist model averaging.

The relatively large proportion of imputed values in some predictor variables (up to 70%) might cause distrust in our results. However, it should be noted that the applied MI procedure completely protects against false inference, as long as the missing data mechanism is correctly modeled and the MAR or MCAR assumptions are met (Enders, 2010). Even a predictor with 90% missing values could still be estimated with MI, albeit with relatively large uncertainty (Steyerberg, 2009). There are several reasons why we believe that the high missing data rate is not a major problem in the present investigation. First, the MAR assumption is highly plausible because missing data almost exclusively resulted from different study designs among the pooled studies. Second, the number of imputed data sets was relatively high, which is recommended with large proportions of missingness (Enders, 2010). Third, even for the predictor with the highest missing data rate (i.e., Absorption), the loss of statistical power induced by missingness was moderate and did not inhibit the detection of statistically significant associations.

5.4.2 *Conclusions*

Although drug dose was clearly the most important determinant of psilocybin response, the results of this study confirm that a substantial proportion of the intra- and interindividual differences in acute responses to psilocybin is related to differences in set and setting. The results suggest that important predictors of psilocybin response can be found in a wide range of different domains, including personality, current mood, psychopathology, drug pre-experience, demography, and environment.

Part III

GENERAL DISCUSSION AND PERSPECTIVES

6 GENERAL DISCUSSION

6.1 TOLERABILITY

The first empirical study of this thesis investigated acute, subacute, and long-term subjective effects of psilocybin in a sample of 110 subjects who had received 1-4 oral doses of psilocybin (45-315 µg/kg body weight) in the context of double-blind, placebo-controlled experimental studies. The primary aim of this study was to report about the subjective tolerability of psilocybin when administered to healthy volunteers in a supervised research environment.

Consistent with recent smaller scale studies (Hasler, Grimberg, et al., 2004; Carhart-Harris, Williams, et al., 2011), psilocybin dose-dependently induced marked alterations in all mental functions, including perception, mood, volition, cognition, and self-experience. Furthermore, dose-response relationships were linear within the administered dose range for all major dimensions of the 5D-ASC questionnaire. Of all main scales of the 5D-ASC, psilocybin most strongly increased the Visionary Restructuralization (VRS) scale followed by the Oceanic Boundlessness (OBN) scale. The Dread of Ego Dissolution (DED) scale, which covers experiences that are characteristic for so called "bad trips", was the least affected 5D-ASC main scale. Furthermore, changes in the DED scale were primarily due to unpleasant disturbances of cognitive functions and somatesthesia, and much less so to suspiciousness or paranoid ideation. Very high DED scores (> 70% of the possible maximum) were only measured in the 250-260 µg/kg and 315 µg/kg dose conditions in small percentages of subjects (i.e. 5.7% and 7.3%, respectively). However, all of these acute adverse reactions were successfully managed by providing interpersonal support and reassurance (e.g. by touching the arm or shoulder with verbal reminders that the participant is in a research study, has taken the hallucinogen, and that he or she will return to normal consciousness). Furthermore, it was never necessary to administer a tranquilizer. In accordance with a study of Heimann (1961), psilocybin effects as measured by the EWL-60-S subscales differed in their time-courses. Whereas the effects on emotional excitation, sensitivity, heightened mood, and concentration reached their maximum in an early phase (60-180 min after drug intake), the effects on dreaminess, dazed state, inactivation, and introversion were more pronounced in a later phase (260-400 min).

Measures of mood and somatic and psychological ailments indicated that subacute side effects of psilocybin were generally mild and that usual functioning was almost completely restored 24 h after drug administration. The most frequent subacute side effects of psilocybin were symptoms of fatigue and headache, which occurred in 60% and 37.5% of subjects, respectively, in the highest dose condition. These findings are in agreement with an earlier study of Hollister (1961), which also found occasional headaches and fatigue as being the most frequent after effects of psilocybin, and provide further evidence that psilocybin is physiologically well tolerated (see also Passie, Seifert, et al., 2002).

A follow-up questionnaire that contained items about (1) acute drug experiences in retrospect, (2) changes in values and attitudes, (3) changes in drug consumption habits, (4) spontaneously occurring ASCs before and after the experiments and flashbacks, and (5) negative changes in psychological well-being and/or mental functions was completed by 90 of the 110 subjects 8-16 months after the last experimental session.

Relatively large numbers of subjects rated the acute drug experience retrospectively as very enriching and influential and only few subjects as very frightening and unpleasant. These results are consistent with recent studies of Griffiths and his colleagues (Griffiths, Richards, Johnson, et al., 2008; Griffiths, Johnson, et al., 2011), which demonstrated that psilocybin can facilitate experiences having enduring personal meaning and spiritual significance. Interestingly, almost all subjects who experienced strong DED during the session evaluated the experience as enriching in retrospect, which suggests that unpleasant and anxious reactions to psilocybin were mostly positively integrated in the long run. Similar results have been recently reported by Griffiths, Johnson, et al. (2011). They found that psychological struggle during the session did neither affect the rate of mystical experiences nor subsequent ratings of the session as having personal meaning and spiritual significance.

Significant numbers of subjects reported psilocybin-induced positive changes in attitudes at the long-term follow-up. The most frequently reported positive changes concerned attitudes toward ASCs (56% of subjects), environment/nature (38%), and aesthetic experiencing (37%). However, even though similar changes have been found in earlier follow up studies (McGlothlin, Cohen, & McGlothlin, 1967; Doblin, 1991; Griffiths, Richards, Johnson, et al., 2008), these results must be interpreted cautiously because they have neither been validated on behavioral measures nor on information provided by close relatives and friends. Hence, they could be biased towards preconceived opinions and expectations of subjects.

Only few subjects reported changes in drug consumption habits, and those who did more often described decreased than increased substance consumption. Furthermore, there was no indication that any subject developed an abusive pattern of drug use. These results are in accordance with a large number of animal (e.g., Fantegrossi et al., 2008) and human studies (e.g., McGlothlin & Arnold, 1971), which have consistently shown that classical hallucinogens are not addictive.

Detailed questions about possible flashback phenomena and spontaneous ASC in the follow-up questionnaire indicated that no subject had experienced Hallucinogen Persisting Perception Disorder (HPPD) or flashbacks as defined in the DSM-IV and ICD-10, respectively, subsequent to the experiments. This supports the view that HPPD and other troubling perceptual abnormalities rarely occur in a therapeutic or research context, where judicious doses of pharmaceutical quality drugs are given (Strassman, 1984; Halpern & Pope, 2003). Apart from the research setting, the use of psilocybin might also have contributed to absence of HPPD and flashbacks. Psilocybin has a relatively short duration of action and therefore seems to produce flashbacks less frequently than LSD (Hermle, Kovar, et al., 2008).

The follow-up questionnaire also provided no indication that any of the subjects had experienced prolonged psychosis or depression. However, one of the 110 subjects experienced symptoms of emotional instability, anxiety, and depression in the first few weeks after drug administration, which were severe enough for him to seek

professional help. Although the symptoms were relatively benign and completely resolved after a few psychotherapeutic sessions, this case underlines the importance of careful debriefing and follow-up of subjects – especially in the first few days and weeks after drug administration.

Taken together, the study presented in Chapter 3 provided considerable evidence that psilocybin is well tolerated both acutely and in the long run. It is likely that the careful selection, preparation, and monitoring of subjects, as well as the administration of predominantly moderate drug doses, have largely contributed to these results. Hence, the results could be used to justify controlled experiments with these drugs, but should not be generalized to unsupervised and recreational drug use.

6.2 ASSESSMENT

The second empirical study presented in Chapter 4 evaluated the psychometric properties of the OAV, which is one of the most widely used self-report questionnaires for assessing hallucinogen-induced ASCs. The development of the OAV was driven by the hypothesis that ASCs – independent of their means of induction – have features in common that can be parsimoniously described on three etiology independent dimensions: Oceanic Boundlessness (OBN), Dread of Ego Dissolution (DED), and Visionary Restructuralization (VRS; Dittrich, 1985; Bodmer, 1989). Hence, an important goal of this study was to test whether the hypothesized three-dimensional structure was superior to other solutions and whether this structure was invariant across stimulus conditions. Unlike in previous studies, these hypotheses were tested by using methods of structural equation modeling (SEM). SEM methodology is considered more appropriate for testing specific hypotheses on latent structures than the previously used exploratory methods (Brown, 2006). The study also used a larger sample than previous studies. Specifically, it was based on a sample of 591 measures of ASC, 327, 162, and 102 of which were experimentally induced by psilocybin, ketamine, and MDMA, respectively.

The originally proposed three-dimensional structure provided a poor fit to the data when modeled with a simple structure CFA. However, because such models are sometimes unnecessarily restrictive (Marsh, Lüdtke, et al., 2010), several attempts were made to model it in a less restrictive way: (1) by freeing residual correlations (i.e. introducing method effects), (2) by freeing cross-loadings (i.e. relaxing the simple structure assumption), and (3) by introducing an additional general factor (i.e. modeling it as a bifactor model). Although freeing these restrictions considerably improved model fit, the overall model fit was still unacceptable even in the least restrictive model. This was mainly because all three factors were multidimensional and because the VRS factor contained several items that loaded more strongly on the OBN than on the VRS factor. Interestingly, the latter was also found in both of the two other studies that have reexamined the factorial structure of the OAV (Habermeyer, 1999; Bodmer, 1999).

Since the original model could not be rescued by relaxing assumptions, the hierarchical structure of the OAV was explored by using Revelle's ICLUST algorithm (Revelle, 1978). The analysis revealed that the OBN and VRS factors merged on a high level of the construct hierarchy. However, on a lower level of the hierarchy, the original factors split into 11 homogenous factors, which were named as follows: (1) Experience of Unity, (2) Spiritual Experience, (3) Blissful State, (4) Insightfulness, (5) Disembodiment, (6)

Impaired Control and Cognition, (7) Anxiety, (8) Complex Imagery, (9) Elementary Imagery, (10) Audio-Visual Synesthesiae, and (11) Changed Meaning of Percepts. A simple structure CFA with these 11 factors provided an acceptable fit to the data. Furthermore, MIMIC modeling indicated that these 11 factors were measurement invariant across the three drugs, physical environments of the drug sessions (PET vs. non-PET), questionnaire versions (OAV vs. 5D-ASC), and sexes. Thus, the summated scales of these factors can be used to make unbiased comparisons across these conditions.

Because the new scales were unidimensional whereas the old scales were not, the new scales were more reliable when reliability is defined as the proportion of variance that is due to one general factor. Nevertheless, the original OAV still showed relatively strong general factor saturations. Over 60% of the variance in the total scale and over 70% of the variances in the OBN, DED, and VRS scales were explained by general factors. Thus, although these scales form ambiguous correlations with other psychological constructs, they could still be useful for the prediction of complex criteria (cf. Revelle, 1979; Rossiter, 2002).

Correlations with the EWL-60-S and STAI-S scales indicated that the new OAV scales tended to have better convergent and discriminant validities than the old scales. This is most likely because the new scales are more homogenous than the old scales. The new scales also differentiated better among the three drug groups. For example, specific MDMA and ketamine effects, such as strong positive mood and disembodiment, could be well differentiated by the new scales, but were mixed up with other effects in the old scales.

Taken together, the study presented in Chapter 4 demonstrated that the factorial structure of the OAV items is better represented by 11 lower order factors than by the original factors. The new factors have the potential to considerably improve the assessment of psilocybin-, ketamine-, and MDMA-induced ASCs, because they form less ambiguous correlations with other measures, are easier to interpret, and provide important additional information on more specific aspects of ASCs.

6.3 PREDICTION

The aim of the third empirical study presented in Chapter 5 was to examine the relative importance of 24 variables, including age, sex, years of education, body mass index, personality traits, drug pre-experience, mental state before drug intake, psychological distress, experimental setting, and drug dose, in the prediction of acute subjective effects of psilocybin. The new OAV subscales that were constructed and validated in Chapter 4, as well as the original OAV scales, served as dependent variables in this study. The statistical analysis was based on the pooled data of 23 studies involving 409 psilocybin administrations to 261 healthy volunteers.

Drug dose was clearly the most important predictor because it was always retained in automatic variable selections and because it had by far the largest effect on the overall consciousness alteration as measured by the OAV total scale (i.e., G-ASC). The second most important predictor was the personality trait of absorption because it was highly positively associated with G-ASC, mystical type experiences, and visual effects induced by psilocybin. This finding is in agreement with large number of studies showing that Absorption is associated with differential responsivity to various ASC

induction procedures (Pekala, Wenger, & Levine, 1985; Vaitl et al., 2005). Furthermore, it is consistent with a recent study of (Ott et al., 2005), which indicated that Absorption is associated with the T102C polymorphism and, consequently, the binding potential of the 5-HT_{2A} receptor.

Other personality traits, including Neuroticism, had only a minor influence on the effects of psilocybin. This is rather surprising because personality traits in general and Neuroticism in particular have often been associated with hallucinogenic drug effects (e.g., Kornetsky & Humphries, 1957; Rinkel et al., 1961; Fischer, Marks, et al., 1968; Barr & Langs, 1972; Hemsley & Ward, 1985; Dittrich, 1994; Lienert & Netter, 1996; Bresnick & Levin, 2006). Some authors have even claimed that personality traits are among the most important determinants of drug response (e.g., Fischer, Marks, et al., 1968; Barr & Langs, 1972). One explanation for this discrepancy might be that in this study the regression coefficients were adjusted for potential confounders, whereas in most other studies they were not. Other explanations are differences in measurement instruments, sample characteristics, and setting. For example, due to the nature of the recruitment, the participants of this study had a mean Neuroticism score that was almost one standard deviation below the population average. Furthermore, as demonstrated in the first empirical study of this thesis, acute adverse reactions were relatively rare. Hence, the association between Neuroticism and acute adverse reactions might have been blurred in this sample.

The mood state immediately before drug intake and psychological distress in the past four weeks in general predicted drug effects more strongly than personality traits, which is in agreement with existing studies (Metzner et al., 1965; Dittrich & Lamparter, 1994). Specifically, Performance-related activity was relatively strongly positively associated with G-ASC and with various experiences covered by the OBN and VRS dimensions, whereas Emotional excitability was strongly positively associated with Spiritual Experience and Anxiety and moderately with all OAV main scales, Insightfulness, and Audio-Visual Synesthesiae. By contrast, psychological distress as measured by the total scale of the SCL-90-R was negatively associated with OBN, Blissful State, and Complex Imagery.

Experiments involving PET measurements were strongly positively associated with Anxiety. In fact, the effect of PET on Anxiety was even larger than that of drug dose. One possible explanation for this large effect could be that the atmosphere at the PET center was perceived as cold and clinical, which is well known to increase the likelihood of anxious reactions (see, Johnson, Richards, & Griffiths, 2008).

Age was negatively associated with Impaired Control and Cognition, whereas gender did not significantly predict any outcome. This is consistent with both human hallucinogen studies (Hyde, 1960; Metzner et al., 1965; Leary et al., 1963) and neuroimaging studies, which have found decreasing 5-HT_{2A} receptors densities with increasing age (Adams et al., 2004) and no differences in 5-HT_{2A} receptors between genders (Frokjaer et al., 2008).

There were also moderate associations between drug pre-experience and drug response. Specifically, hallucinogen-naïve subjects tended to report more Disembodiment, VRS, and Changed Meaning of Percepts; subjects who smoked cannabis at least once per month experienced more Blissful state than those who never smoked cannabis; and regular alcohol consumption was positively associated with Audio-Visual Synesthesiae

and Complex Imagery. These results partially confirm those of Metzner et al. (1965), but are contradicting those of Dittrich and Lamparter (1994).

In conclusion, the results of this study confirmed that a large proportion of inter- and intraindividual differences in acute responses to psilocybin is explained by non-pharmacological variables. Moreover, important predictors can be found in a wide range of domains, including personality, mood, drug pre-experience, psychopathology, setting, and demography.

7

PERSPECTIVES

The present thesis investigated the tolerability, assessment, and prediction of subjective psilocybin effects using the pooled the data of controlled experimental studies. All of the studies on which this thesis is based were carried out by Franz Vollenweider's research group at the University Hospital of Psychiatry in Zurich, which was one of the first to get governmental approval for the use psilocybin in human research after the nearly worldwide moratorium for many decades. Because Vollenweider's group has continuously collected data since the 1990s and because regulatory hurdles in most countries are bigger than in Switzerland, the data set on which this thesis is based is unique in the world with respect to its size, recency, and standardization. Hence, the richness of the data is certainly the biggest strength of this thesis. Since the data offered an unique opportunity to answer important research questions, I put a lot of effort to conduct the statistical analyses in accordance with the highest methodological standards. Consequently, I consider this as another major strength of this thesis.

However, there is still much room for improvement. A major problem that concerns almost all human hallucinogen studies is that the effects of hallucinogens could be biased towards expectancies of investigators and participants. The study of McGlothlin and Arnold (1971) indicates that preconceived opinions and expectancies can have a serious impact, especially on reported long-term beneficial effects of these drugs. The standard method for dealing with expectancy effects is to administer the drug in a double-blind placebo-controlled fashion, and this has also been done in almost all studies of Vollenweider's research group. However, means and standard deviations of most 5D-ASC items obtained under placebo conditions were virtually zero, suggesting the effects of psilocybin were so powerful that they could easily be distinguished from placebo by most subjects. Thus, at least with regard to subjective effects, the use of a placebo control seems to be of limited usefulness. The limited control of expectancy effects might have biased the results of all three studies of this thesis to some degree, but the problem was likely most severe in the long-term follow-up measure of the first study presented in Chapter 3 because the follow-up questionnaire was unvalidated and there was no placebo control (i.e. all subjects knew that they had received psilocybin at least once).

To increase the success of the blinding, some investigators have used an active placebo, such as nicotinic acid (e.g., Pahnke, 1963) or methylphenidate (e.g., Griffiths, Richards, McCann, & Jesse, 2006). Griffiths, Richards, McCann, and Jesse (2006) additionally tried to control expectancy effects by studying participants without personal histories of hallucinogen use; randomly assigning participants into two groups with different number, order, and blinding of drug administration; and by using instructional sets that provided the expectation that sessions could involve not only the administration of a wide range of psilocybin doses but also a range of novel drugs, some of which could produce effects that overlap with those produced by psilocybin. They reported that in 23% of all drug sessions at least one of the two monitors did not correctly identify the

administered drug. Furthermore, they observed that in two of the three occasions in which the primary monitors misclassified a high dose psilocybin session as an active placebo session, the participants nevertheless rated their experiences as among the five most meaningful and spiritually significant experiences of their lives, suggesting that these experiences were not merely artifacts of monitor expectation or suggestion. Although Griffiths, Richards, McCann, and Jesse (2006) did not directly ask participants to identify the drug they had received, it seems that subjects were at least partially blind to the drug because some subjects rated their experience as being among the most personally meaningful experiences of their lives even though they had received methylphenidate. On the other hand, this finding could also be taken as an indication for the powerful effects of suggestion that were at work in this study because it is hard to believe that methylphenidate alone could induce experiences that are equally significant as, for example, the birth of a child or a marriage.

Although the methods used by Griffiths, Richards, McCann, and Jesse (2006) seem to be helpful, expectancy effects might be even more rigorously controlled in human hallucinogen research by adopting the so called balanced-placebo design (BPD). The BPD is considered the optimal method for separating pharmacological effects from the cognitive expectations of receiving the drug (Marlatt & Rohsenow, 1980). It is a 2 × 2 factorial design that crosses the administered substance (drug vs. placebo) with an instructional set manipulation (subjects are told they receive the active drug vs. subjects are told they receive placebo). If the instructional set manipulation is credible, this allows a determination of expectancy effects independent of pharmacological effects and vice versa and the "antiplacebo" effect of the pharmacological action of the drug when expecting no drug (Vogel-Sprott & Fillmore, 1999; Metrik et al., 2009). To ensure the credibility of the instructional set manipulation, various techniques have been developed, such as pouring from an unopened liquor bottle and false breath alcohol feedback in the case of alcohol or simulating the smell and appearance of a cigarette in the case of nicotine (Metrik et al., 2009). Although the BPD has not yet been used in experiments with classical hallucinogens, a recent study has demonstrated its feasibility with marijuana (Metrik et al., 2009). It is therefore likely that it would also work in experiments with classical hallucinogens. A direct comparison of various techniques for controlling expectancy effects (e.g., the regular placebo controlled design vs. the methods of Griffiths described above vs. the BPD) could be very useful not only for evaluating their feasibility and efficiency, but also for evaluating how much self-reported psilocybin effects are affected by expectations.

Besides controlling expectancy effects, it would also be helpful to compare self-reported effects of psilocybin with behavioral measures and judgments of external raters. Unfortunately, this could not be done in the present thesis because comparable objective measures were mostly unavailable. By contrast, Griffiths and his colleagues (Griffiths, Richards, McCann, & Jesse, 2006; Griffiths, Johnson, et al., 2011) have used community observers to corroborate potential changes in attitudes and behavior. However, because external raters can be biased by preconceived opinions and expectancies as well, it is highly important to ensure that the external raters are objective. Unfortunately, the objectivity of the community raters in Griffiths studies must be questioned because the community observers were chosen by the participants themselves. Hence, it is likely that they had very similar opinions about hallucinogens as the participants.

Although the study presented in Chapter 4 has contributed to an improved assessment of psilocybin-induced ASCs by constructing and validating 11 new subscales of the OAV questionnaire, the OAV has still several problems that could not be solved in this thesis. For example, the items of the OAV and its revised version 5D-ASC do not sufficiently cover affective states induced by hallucinogens. Thus, in order to assess all important aspects of classical hallucinogens, they still have to supplemented by other rating scales, such as the EWL-60-S (Janke & Debus, 1986), the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), or the Profile of Mood Scale (POMS; McNair, Lorr, & Droppleman, 1971).

Another problem of the OAV and 5D-ASC is that their item response format is strictly unipolar. As a result, hallucinogen-induced changes assessed by these questionnaires can only occur in one direction. For most of the changes, this is not a problem because changes in the opposite direction can not occur. However, some changes, such as for example changes in anxiety, could theoretically occur in both directions and thus are censored when assessed by unipolar items. Since only changes in the positive direction are assessed and changes in the negative direction are censored, it has occurred that psilocybin on average slightly decreased anxiety when assessed by the STAI-S, but clearly increased anxiety when assessed by the 5D-ASC (M. Kometer, personal communication, August 19, 2011). The unipolar item response format of the OAV and 5D-ASC also leads to strongly skewed item distributions with a large piling up of values at the lower end of the scale. As described in more detail in Chapter 4, such distributions usually can not be transformed to normality with non-linear transformations and can severely bias model parameters and goodness of fit of structural equation models.

Because the above problems are difficult to solve by modifying existing questionnaires, it would be worthwhile to construct a new questionnaire assessing hallucinogen-induced ASCs. To improve on existing measures, the item selection should be guided by theory as much as possible, the item response format should be bipolar and result in normally distributed scores, and the item anchors should be clearly defined such that they are not subject to individual interpretation.

Part IV

APPENDIX

BIBLIOGRAPHY

Abraham, H. D. (1982). A chronic impairment of colour vision in users of LSD. *British Journal of Psychiatry*, *140*, 518–520. doi:10.1192/bjp.140.5.518

Abraham, H. D. (1983). Visual phenomenology of the LSD flashback. *Archives of General Psychiatry*, *40*(8), 884–889. doi:10.1001/archpsyc.1983.01790070074009

Abraham, H. D., & Aldridge, A. M. (1993). Adverse consequences of lysergic acid diethylamide. *Addiction*, *88*(10), 1327–1334. doi:10.1111/j.1360-0443.1993.tb02018.x

Abraham, H. D., & Duffy, F. H. (1996). Stable quantitative EEG difference in post-LSD visual disorder by split-half analysis: Evidence for disinhibition. *Psychiatry Research: Neuroimaging*, *67*(3), 173–187. doi:10.1016/0925-4927(96)02833-8

Abraham, H. D., & Duffy, F. H. (2001). EEG coherence in post-LSD visual hallucinations. *Psychiatry Research: Neuroimaging*, *107*(3), 151–163. doi:10.1016/S0925-4927(01)00098-1

Abraham, H. D., & Wolf, E. (1988). Visual function in past users of LSD: Psychophysical findings. *Journal of Abnormal Psychology*, *97*(4), 443–447. doi:10.1037//0021-843X.97.4.443

Adams, K. H., Pinborg, L. H., Svarer, C., Hasselbalch, S. G., Holm, S., Haugbøl, S., … Knudsen, G. M. (2004). A database of [^{18}F]-altanserin binding to 5-HT$_{2A}$ receptors in normal volunteers: Normative data and relationship to physiological and demographic variables. *Neuroimage*, *21*(3), 1105–1113. doi:10.1016/j.neuroimage.2003.10.046

American Psychological Association. (2010). *Publication manual of the american psychological association* (6. ed., 1. printing). Washington, DC: American Psychological Association.

Ametamey, S., Vollenweider, F. X., Patt, J., Bourquin, D., Hasler, F., Beer, H. F., & Schubiger, P. A. (1998). ^{11}C-radiolabeling of hallucinogenic psilocin, a potential radioligand for studying the role of serotonin receptors in psychotic symptom formation. *Journal of Labelled Compounds & Radiopharmaceuticals*, *41*(7), 585–594. doi:10.1002/(SICI)1099-1344(199807)41:7<585::AID-JLCR115>3.0.CO;2-U

Anderson, M. C., & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science*, *18*(4), 189–194. doi:10.1111/j.1467-8721.2009.01634.x

Angleitner, A., Riemann, R., & Spinath, F. M. (2004). Investigating the ZKPQ-III-R: Psychometric properties, relations to the five-factor model, and genetic and environmental influences in its scales and facets. In R. M. Stelmack (Ed.), *On the psychobiology of personality: Essays in honor of Marvin Zuckerman* (pp. 89–105). Sand Diego, CA: Elsevier.

Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*(3), 397–438. doi:10.1080/10705510903008204

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Bacon, D. R. (2001). An evaluation of cluster analytic approaches to initial model specification. *Structural Equation Modeling*, *8*(3), 397–429. doi:10.1207/S15328007SEM0803_4

Baggott, M. J., Siegrist, J. D., Galloway, G. P., Robertson, L. C., Coyle, J. R., & Mendelson, J. E. (2010). Investigating the mechanisms of hallucinogen-induced visions using 3,4-methylenedioxyamphetamine (MDA): A randomized controlled trial in humans. *PLoS ONE*, *5*(12), e14074. doi:10.1371/journal.pone.0014074

Barrett, P. T. (2006). *Orthosim* [Computer software]. Version 2.1. Retrieved from http://www.pbarrett.net/orthosim2.htm

Barr, H. L., & Langs, R. J. (1972). *LSD: Personality and experience* (I. B. Weiner, Ed.). New York, NY: Wiley-Interscience.

Baumann, U., & Dittrich, A. (1975). Konstruktion einer deutschsprachigen Psychotizismus-Skala [Construction of a German language psychoticism scale]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *22*(3), 421–443.

Baumann, U., & Dittrich, A. (1976). Überprüfung der Fragebogendimension P (Psychotizismus) im Vergleich zur Extraversion und Neurotizismus [Assessment of the questionnaire dimension P (Psychoticism) in comparison with Extraversion and Neuroticism]. *Zeitschrift für Klinische Psychologie – Forschung und Praxis*, *5*(1), 1–23.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Bickel, P., Dittrich, A., & Schöpf, J. (1976). Eine experimentelle Untersuchung zur bewusstseinsverändernden Wirkung von N,N-Dimethyltryptamin (DMT) [An experimental study on altered states of consciousness induced by N,N-dimethyltryptamine (DMT)]. *Pharmakopsychiatry*, *9*(5), 220–225. doi:10.1055/s-0028-1094495

Blanchard-Fields, F. (2007). Everyday problem solving and emotion. *Current Directions in Psychological Science*, *16*(1), 26–31. doi:10.1111/j.1467-8721.2007.00469.x

Bodmer, I. (1989). *Konstruktion des Fragebogens OAV zur quantitativen Erfassung aussergewöhnlicher Bewusstseinszustände (ABZ) [Construction of the OAV questionnaire for quantitatively assessing altered states of consciousness]*. (Unpublished master's thesis, University of Zurich, Switzerland).

Bodmer, I. (1999). *Erinnerung an einen aussergewöhnlichen Bewusstseinszustand: Eine experimentelle Untersuchung zum autobiographischen Gedächtnis [Recollection of an altered state of consciousness: An experimental study on the autobiographic memory]*. Berlin, Germany: VWB.

Bodmer, I., Dittrich, A., & Lamparter, D. (1994). Aussergewöhnliche Bewusstseinszustände - Ihre gemeinsame Struktur und Messung [Altered states of consciousness - Their common structure and assessment]. In A. Hofmann & H. Leuner (Eds.), *Welten des Bewusstseins. Bd. 3: Experimentelle Psychologie, Neurobiologie und Chemie [Worlds of Consciousness. Vol. 3 - Experimental Psychology, Neurobiology, and Chemistry]* (pp. 45–58). Berlin, Germany: VWB.

Bolle, R. H. (1988). *Am Ursprung der Sehnsucht: Tiefenpsychologische Aspekte veränderter Wachbewusstseinszustände am Beispiel des Anästhetikums KETANEST [At the origin of the desire: Depth-psychological aspects of altered states of waking consciousness on the example of the anesthetic KETANEST]*. Berlin, Germany: VWB.

Braun, I. (1997). *Zur quantitativen Beschreibung aussergewöhnlicher Bewusstseinszustände im fünfdimensionalen Raum: Eine empirische Studie [On the quantitative description of altered states of consciousness in the five-dimensional space: An empirical study]*. (Unpublished master's thesis, University of Zurich, Switzerland).

Bresnick, T., & Levin, R. (2006). Phenomenal qualities of ayahuasca ingestion and its relation to fringe consciousness and personality. *Journal of Consciousness Studies*, *13*(9), 5–24.

Bringhurst, R. (2002). *The elements of typographic style*. Version 2.5. Point Roberts, WA: Hartley & Marks.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Methodology in the social sciences. New York, NY: Guilford.

Burt, C. L. (1948). The factorial study of temperamental traits. *British Journal of Psychology – Statistical Section*, *1*, 178–203.

Byrne, B. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. (2nd). London: Taylor and Francis.

Carhart-Harris, R. L., Leech, R., Williams, T. M., Erritzoe, D., Abbasi, N., Bargiotas, T., … Nutt, D. J. (2012). Implications for psychedelic-assisted psychotherapy: a functional magnetic resonance imaging study with psilocybin. *British Journal of Psychiatry*. doi:10.1192/bjp.bp.111.103309

Carhart-Harris, R. L., & Nutt, D. J. (2010). User perceptions of the benefits and harms of hallucinogenic drug use: A web-based questionnaire study. *Journal of Substance Use*, *15*(4), 283–300. doi:10.3109/14659890903271624

Carhart-Harris, R. L., Williams, T. M., Sessa, B., Tyacke, R. J., Rich, A. S., Feilding, A., & Nutt, D. J. (2011). The administration of psilocybin to healthy, hallucinogen-experienced volunteers in a mock-functional magnetic resonance imaging environment: A preliminary investigation of tolerability. *Journal of Psychopharmacology*, *25*(11), 1562–1567. doi:10.1177/0269881110367445

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, *10*(1), 1–19. doi:10.1007/BF02289789

Carter, O. L., Burr, D. C., Pettigrew, J. D., Wallis, G. M., Hasler, F., & Vollenweider, F. X. (2005). Using psilocybin to investigate the relationship between attention, working memory, and the serotonin 1A and 2A receptors. *Journal of Cognitive Neuroscience*, *17*(10), 1497–1508. doi:10.1162/089892905774597191

Carter, O. L., Hasler, F., Pettigrew, J. D., Wallis, G. M., Liu, G. B., & Vollenweider, F. X. (2007). Psilocybin links binocular rivalry switch rate to attention and subjective arousal levels in humans. *Psychopharmacology (Berlin)*, *195*(3), 415–424. doi:10.1007/s00213-007-0930-9

Carter, O. L., Pettigrew, J. D., Burr, D. C., Alais, D., Hasler, F., & Vollenweider, F. X. (2004). Psilocybin impairs high-level but not low-level motion perception. *Neuroreport*, *15*(12), 1947–1951. Retrieved from http://journals.lww.com/neuroreport/Abstract/2004/08260/Psilocybin_impairs_high_level_but_not_low_level.23.aspx

Carter, O. L., Pettigrew, J. D., Hasler, F., Wallis, G. M., Liu, G. B., Hell, D., & Vollenweider, F. X. (2005). Modulating the rate and rhythmicity of perceptual rivalry alternations with the mixed 5-HT_{2A} and 5-HT_{1A} agonist psilocybin. *Neuropsychopharmacology*, *30*(6), 1154–1162. doi:10.1038/sj.npp.1300621

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. doi:10.1207/s15327906mbr0102_10

Cerletti, A. (1959). Pharmacology of psilocybin. In P. B. Bradley, P. Deniker & C. Radouco-Thomas (Eds.), *Neuro-pharmacology* (1st ed., pp. 291–294). Amsterdam, The Netherlands: Elsevier.

Chan, W., Ho, R. M., Leung, K., Chan, D. K., & Yiu-Fai, Y. (1999). An alternative method for evaluating congruence coefficients with Procrustes rotation: A bootstrap procedure. *Psychological Methods*, *4*(4), 378–402. doi:10.1037/1082-989X.4.4.378

Clark, B. (1968). Some early observations on the use of psiloybin in psychiatric patients. *British Journal of Social Psychiatry*, *2*, 21–25. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=5654

Cohen, S. (1960). Lysergic acid diethylamide: Side effects and complications. *Journal of Nervous and Mental Disease*, *130*, 30–40. Retrieved from http://journals.lww.com/jonmd/Citation/1960/01000/Lysergic_Acid_Diethylamide__Side_Effects_and.5.aspx

Cohen, S., & Ditman, K. S. (1963). Prolonged adverse reactions to lysergic acid diethylamide. *Archives of General Psychiatry*, *8*, 475–480. doi:10.1001/archpsyc.1963.01720110051006

Cooksey, R. W., & Soutar, G. N. (2006). Coefficient beta and hierarchical item clustering: An analytical procedure for establishing and displaying the dimensionality and homogeneity of summated scales. *Organizational Research Methods*, *9*(1), 78–98. doi:10.1177/1094428105283939

Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. Boyle, G. Matthews & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (2nd ed., pp. 179–198). Thousand Oaks, CA: SAGE.

Crawford, A. V., Green, S. B., Levy, R., Lo, W., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885–901. doi:10.1177/0013164410379332

Csomor, P. A., Yee, B. K., Feldon, J., Theodoridou, A., Studerus, E., & Vollenweider, F. X. (2009). Impaired prepulse inhibition and prepulse-elicited reactivity but intact reflex circuit excitability in unmedicated schizophrenia patients: A comparison with healthy subjects and medicated schizophrenia patients. *Schizophrenia Bulletin*, *35*(1), 244–255. doi:10.1093/schbul/sbm146

Daumann, J., Heekeren, K., Neukirch, A., Thiel, C. M., Möller-Hartmann, W., & Gouzoulis-Mayfrank, E. (2008). Pharmacological modulation of the neural basis underlying inhibition of return (IOR) in the human 5-HT_{2A} agonist and NMDA antagonist model of psychosis. *Psychopharmacology (Berlin)*, *200*(4), 573–583. doi:10.1007/s00213-008-1237-1

Daumann, J., Wagner, D., Heekeren, K., Neukirch, A., Thiel, C. M., & Gouzoulis-Mayfrank, E. (2010). Neuronal correlates of visual and auditory alertness in the DMT and ketamine model of psychosis. *Journal of Psychopharmacology*, *24*(10), 1515–1524. doi:10.1177/0269881109103227

David, A. E., & David, J. M. (1961). La psilocibina, un nuevo alucinógeno, y sus posibilidades terapéuticas en psicoterapia [Psilocybin, a new hallucinogen, and its therapeutic possibilities in psychotherapy]. *Acta Neuropsiquiatrica Argentina*, *7*,

143–144. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=2636

Deneau, G., Yanagita, T., & Seevers, M. H. (1969). Self-administration of psychoactive substances by the monkey. *Psychopharmacologia*, *16*(1), 30–48. doi:10.1007/BF00405254

Derogatis, L. R. (1994). *SCL-90-R: Symptom Checklist-90-R. Administration, scoring and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.

Diezi, P., Faeh, M., & Hermann, E. (1982). *Die Prognostizierbarkeit visuell-halluzinatorischer Phänomene [The predictability of visual hallucinatory phenomena]*. (Unpublished master's thesis, University of Zurich, Switzerland).

Dittrich, A. (1975a). Vergleich veränderter Bewusstseinszustände induziert durch kurzandauernden Reizentzug und durch (–)–Δ^9-trans-Tetrahydrocannabinol. [Comparison of altered states of consciousness induced by short-term sensory deprivation and (–)–Δ^9-trans-tetrahydrocannabinol]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *22*(4), 547–560.

Dittrich, A. (1975b). Zusammenstellung eines Fragebogens (APZ) zur Erfassung abnormer psychischer Zustände [Construction of a questionnaire (APZ) for assessing abnormal mental states]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *23*(1), 12–20.

Dittrich, A. (1985). *Ätiologie-unabhängige Strukturen veränderter Wachbewusstseinszustände. Ergebnisse empirischer Untersuchungen über Halluzinogene I. und II. Ordnung, sensorische Deprivation, hypnagoge Zustände, hypnotische Verfahren sowie Reizüberflutung [Etiology-independent structures of altered states of consciousness. Results of empirical studies on hallucinogens of the first and second order, sensory deprivation, hypnagogic states, hypnotic procedures, and sensory overload]*. Stuttgart, Germany: Enke.

Dittrich, A. (1994). Psychological aspects of altered states of consciousness of the LSD type: Measurements of their basic dimensions and prediction of individual differences. In A. Pletscher & D. Ladewig (Eds.), *50 years of LSD. Current status and perspectives of hallucinogens* (pp. 101–118). New York, NY: Parthenon.

Dittrich, A. (1996). *Ätiologie-unabhängige Strukturen veränderter Wachbewusstseinszustände. Ergebnisse empirischer Untersuchungen über Halluzinogene I. und II. Ordnung, sensorische Deprivation, hypnagoge Zustände, hypnotische Verfahren sowie Reizüberflutung [Etiology-independent structures of altered states of consciousness. Results of empirical studies on hallucinogens of the first and second order, sensory deprivation, hypnagogic states, hypnotic procedures, and sensory overload]* (2nd ed.). Berlin, Germany: VWB.

Dittrich, A. (1998). The standardized psychometric assessment of altered states of consciousness (ASCs) in humans. *Pharmacopsychiatry*, *31*, 80–84. doi:10.1055/s-2007-979351

Dittrich, A., Baettig, K., Woggon, B., & Zeppelin, I. V. (1972). Entwicklung einer Selbsteinschätzungsskala (DAE-Skala I) zur Erfassung von Cannabis-Effekten [Development of an auto-estimation scale (DAE-Scale I) for determining the effects of cannabis]. *Pharmacopsychiatry*, *5*(5), 255–268. doi:10.1055/s-0028-1094355

Dittrich, A., Bickel, P., Schöpf, J., & Zimmer, D. (1976). Vergleich veränderter Bewusstseinszustände unter den Halluzinogenen (–)– Δ^9-trans-Tetrahydrocannabinol (Δ^9-THC) und *N,N*-Dimethyltryptamin [Comparison of altered states of consciousness induced by the hallucinogens (–)–Δ^9-trans-tetrahydrocannabinol (Δ^9-THC)

and *N,N*-dimethyltryptamine (DMT)]. *Archiv für Psychiatrie und Nervenkrankheiten*, *223*(1), 77–87.

Dittrich, A., Bickel, P., & Zimmer, D. (1975). Effekte von (–)–Δ^9-trans-Tetrahydrocannabinol (Δ^9-THC) auf Psychotizismus-Tests. Eine Untersuchung des Eysenckschen Drogenpostulats über Beziehungen zwischen Halluzinogenwirkung und Psychotizismus [Effects of (–)–Δ^9-trans-tetrahydrocannabinol (Δ^9-THC) on tests of psychoticism. A study of Eysenck's drug postulate on the effects of hallucinogens in relation to psychoticism]. *Psychopharmacologia*, *40*(4), 351–358. doi:10.1007/BF00421474

Dittrich, A., & Lamparter, D. (1994). Differentielle Psychologie aussergewöhnlicher Bewusstseinszustände (ABZ) - Ergebnisse experimenteller Untersuchungen mit sensorischer Deprivation, N,N-Dimethyltryptamin und Stickoxydul [Differential Psychology of altered states of consciousness (ASC) - Results of experimental studies with sensory deprivation, N,N-dimethyltryptamine, and nitrous oxide]. In A. Hofmann & H. Leuner (Eds.), *Welten des Bewusstseins. Bd. 3: Experimentelle Psychologie, Neurobiologie und Chemie [Worlds of Consciousness. Vol. 3 - Experimental Psychology, Neurobiology, and Chemistry]* (pp. 71–86). Berlin, Germany: VWB.

Dittrich, A., von Arx, S., & Staub, S. (1985). International study on altered states of consciousness (ISASC) – Summary of the results. *The German Journal of Psychology*, *9*(4), 319–339.

Dittrich, A., Lamparter, D., & Maurer, M. (1989). *Itemsammlung zum Fragebogen BETA [Item collection for the BETA questionnaire]*. Unpublished manuscript.

Dittrich, A., Lamparter, D., & Maurer, M. (1999). *5D-ABZ: Fragebogen zur Erfassung Aussergewöhnlicher Bewusstseinszustände. Eine kurze Einführung [5D-ASC: Questionnaire for the assessment of altered states of consciousness. A short introduction]* (1st ed.). Zurich, Switzerland: PSIN PLUS.

Dittrich, A., Lamparter, D., & Maurer, M. (2006). *5D-ABZ: Fragebogen zur Erfassung Aussergewöhnlicher Bewusstseinszustände. Eine kurze Einführung [5D-ASC: Questionnaire for the assessment of altered states of consciousness. A short introduction]* (2nd ed.). Zurich, Switzerland: PSIN PLUS.

Dittrich, A., Lamparter, D., & Maurer, M. (2010). *5D-ASC: Questionnaire for the assessment of altered states of consciousness. A short introduction* (3rd ed.). Zurich, Switzerland: PSIN PLUS.

Doblin, R. (1991). Pahnke's Good Friday experiment: A long-term follow-up and methodological critique. *Journal of Transpersonal Psychology*, *23*, 1–28. Retrieved from http://www.maps.org/research/cluster/psilo-lsd/goodfriday.pdf

Duhachek, A., & Lacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, *89*(5), 792–808. doi:10.1037/0021-9010.89.5.792

Dunsmore, M. W., & Kaplan, H. B. (1997). Peer support, perceptions of adverse effects, and the hallucinogenic drug experience. *Applied Behavioral Science Review*, *5*(2), 219–230. doi:10.1016/S1068-8595(97)80016-5

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, *27*(29), 6137–6157. doi:10.1002/sim.3429

Eisner, B. (1997). Set, setting, and matrix. *Journal of Psychoactive Drugs*, *29*(2), 213–216. Retrieved from http://www.erowid.org/references/texts/show/1306docid6091

Enders, C. K. (2010). *Applied missing data analysis*. Methodology in the social sciences. New York, NY: Guilford.

Endler, N. S., Cox, B. J., Parker, J. D., & Bagby, R. M. (1992). Self-reports of depression and state-trait anxiety: Evidence for differential assessment. *Journal of Personality and Social Psychology*, *63*(5), 832–838. doi:10.1037/0022-3514.63.5.832

Espiard, M.-L., Lecardeur, L., Abadie, P., Halbecq, I., & Dollfus, S. (2005). Hallucinogen persisting perception disorder after psilocybin consumption: A case study. *European Psychiatry*, *20*(5-6), 458–460. doi:10.1016/j.eurpsy.2005.04.008

Fachner, J., & Rittner, S. (2004). Sound and trance in a ritualistic setting visualised with EEG brainmapping. *Music Therapy Today*, *5*(2), 1–32. Retrieved from http://www.wfmt.info/Musictherapyworld/modules/mmmagazine/issues/20040225084519/20040225085618/MTT_5_1_FachRitt.pdf

Fahrenberg, J., Hampel, R., & Selg, H. (1978). *Das Freiburger Persönlichkeitsinventar – FPI [The Freiburg Personality Inventory – FPI]* (3rd ed.). Göttingen, Germany: Hogrefe.

Fahrenberg, J., Hampel, R., & Selg, H. (1984). *Das Freiburger Persönlichkeitsinventar – FPI [The Freiburg Personality Inventory – FPI]* (4th ed.). Göttingen, Germany: Hogrefe.

Fantegrossi, W. E., Murnane, K. S., & Reissig, C. J. (2008). The behavioral pharmacology of hallucinogens. *Biochemical Pharmacology*, *75*(1), 17–33. doi:10.1016/j.bcp.2007.07.018

Fischer, R. (1971). A cartography of the ecstatic and meditative states. *Science*, *174*(12), 897–904. doi:10.1126/science.174.4012.897

Fischer, R., Hill, R. M., Thatcher, K., & Scheib, J. (1970). Psilocybin-induced contraction of nearby visual space. *Agents Actions*, *1*(4), 190–197. doi:10.1007/BF01965761

Fischer, R., Hill, R. M., & Warshay, D. (1969). Effects of psychodysleptic drug psilocybin on visual perception. Changes in brightness preference. *Experientia*, *25*(2), 166–169. doi:10.1007/BF01899102

Fischer, R., Marks, P. A., Hill, R. M., & Rockey, M. A. (1968). Personality structure as the main determinant of drug induced (model) psychoses. *Nature*, *218*(5138), 296–298. doi:10.1038/218296a0

Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology: Social Sciences*, *57*(5), 275–284. doi:10.1093/geronb/57.5.S275

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. doi:10.1037/1082-989X.9.4.466

Franke, G. H. (1995). *SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version. [SCL-90-R. The Symptom-Checklist of Derogatis - German version]*. Göttingen, Germany: Beltz.

Frecska, E., & Luna, L. E. (2006). The adverse effects of hallucinogens from intramural perspective. *Neuropsychopharmacologia Hungarica*, *8*(4), 189–200. Retrieved from http://www.mppt.hu/images/magazin/pdf/viii-evfevfolyam-4-szam/a-hallucinogenek-karos-hatasai-terapias-alkalmazasuk.pdf

Frei, E., Gamma, A., Pascual-Marqui, R., Lehmann, D., Hell, D., & Vollenweider, F. X. (2001). Localization of MDMA-induced brain activity in healthy volunteers us-

ing low resolution brain electromagnetic tomography (LORETA). *Human Brain Mapping*, *14*(3), 152–165. doi:10.1002/hbm.1049

Freud, S. (1961). Civilization and its discontents. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press. (Original work published 1930).

Frokjaer, V. G., Mortensen, E. L., Nielsen, F. A., Haugbol, S., Pinborg, L. H., Adams, K. H., … Knudsen, G. M. (2008). Frontolimbic serotonin 2A receptor binding in healthy subjects is associated with personality risk factors for affective disorder. *Biological Psychiatry*, *63*(6), 569–576. doi:10.1016/j.biopsych.2007.07.009

Gamma, A., Buck, A., Berthold, T., Liechti, M. E., Vollenweider, F. X., & Hell, D. (2000). 3,4-Methylenedioxymethamphetamine (MDMA) modulates cortical and limbic brain activity as measured by [$H_2{}^{15}O$]-PET in healthy humans. *Neuropsychopharmacology*, *23*(4), 388–395. doi:10.1016/S0893-133X(00)00130-5

Gelman, A. (2004). *Bayesian data analysis*. Roca Raton, GL: Chapman & Hall.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*(15), 2865–2873. doi:10.1002/sim.3107

Geyer, M. A., & Vollenweider, F. X. (2008). Serotonin research: Contributions to understanding psychoses. *Trends in Pharmacological Sciences*, *29*(9), 445–453. doi:10.1016/j.tips.2008.06.006

Goldberg, L. R. (1981). Language and individuell differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (2nd, pp. 141–165). Beverly Hills, CA: SAGE.

Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S. Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed., pp. 209–237). New York, NY: Springer.

González, D., Riba, J., Bouso, J. C., Gómez-Jarabo, G., & Barbanoj, M. J. (2006). Pattern of use and subjective effects of Salvia divinorum among recreational users. *Drug and Alcohol Dependence*, *85*(2), 157–162. doi:10.1016/j.drugalcdep.2006.04.001

González-Maeso, J., Ang, R. L., Yuen, T., Chan, P., Weisstaub, N. V., López-Giménez, J. F., … Sealfon, S. C. (2008). Identification of a serotonin/glutamate receptor complex implicated in psychosis. *Nature*, *452*(7183), 93–97. doi:10.1038/nature06612

González-Maeso, J., & Sealfon, S. C. (2009). Agonist-trafficking and hallucinogens. *Current Medicinal Chemistry*, *16*(8), 1017–1027. doi:10.2174/092986709787581851

González-Maeso, J., Weisstaub, N. V., Zhou, M., Chan, P., Ivic, L., Ang, R. L., … Gingrich, J. A. (2007). Hallucinogens recruit specific cortical 5-HT_{2A} receptor-mediated signaling pathways to affect behavior. *Neuron*, *53*(3), 439–452. doi:10.1016/j.neuron.2007.01.008

González-Maeso, J., & Sealfon, S. (2009). Psychedelics and schizophrenia. *Trends in Neurosciences*, *32*(4), 225–232. doi:10.1016/j.tins.2008.12.005

Gouzoulis-Mayfrank, E., Habermeyer, E., Hermle, L., Steinmeyer, A., Kunert, H., & Sass, H. (1998). Hallucinogenic drug induced states resemble acute endogenous psychoses: Results of an empirical study. *European Psychiatry*, *13*(8), 399–406. doi:10.1016/S0924-9338(99)80686-5

Gouzoulis-Mayfrank, E., Heekeren, K., Neukirch, A., Stoll, M., Stock, C., Obradovic, M., & Kovar, K. A. (2005). Psychological effects of (S)-ketamine and N,N-dime-

thyltryptamine (DMT): A double-blind, cross-over study in healthy volunteers. *Pharmacopsychiatry*, *38*(6), 301–311. doi:10.1055/s-2005-916185

Gouzoulis-Mayfrank, E., Heekeren, K., Thelen, B., Lindenblatt, H., Kovar, K. A., Sass, H., & Geyer, M. A. (1998). Effects of the hallucinogen psilocybin on habituation and prepulse inhibition of the startle reflex in humans. *Behavioural Pharmacology*, *9*(7), 561–566. Retrieved from http://journals.lww.com/behaviouralpharm/Abstract/1998/11000/Effects_of_the_hallucinogen_psilocybin_on.11.aspx

Gouzoulis-Mayfrank, E., Schneider, F., Friedrich, J., Spitzer, M., Thelen, B., & Sass, H. (1998). Methodological issues of human experimental research with hallucinogens. *Pharmacopsychiatry*, *31 Suppl 2*, 114–118. doi:10.1055/s-2007-979356

Gouzoulis-Mayfrank, E., Schreckenberger, M., Sabri, O., Arning, C., Thelen, B., Spitzer, M., … Sass, H. (1999). Neurometabolic effects of psilocybin, 3,4-methylenedioxyethylamphetamine (MDE) and d-methamphetamine in healthy volunteers. A double-blind, placebo-controlled PET study with [^{18}F]FDG. *Neuropsychopharmacology*, *20*(6), 565–581. doi:10.1016/S0893-133X(98)00089-X

Gouzoulis-Mayfrank, E., Thelen, B., Habermeyer, E., Kunert, H. J., Kovar, K. A., Lindenblatt, H., … Sass, H. (1999). Psychopathological, neuroendocrine and autonomic effects of 3,4-methylenedioxyethylamphetamine (MDE), psilocybin and d-methamphetamine in healthy volunteers. Results of an experimental double-blind placebo-controlled study. *Psychopharmacology (Berlin)*, *142*(1), 41–50. doi:10.1007/s002130050860

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213. doi:10.1007/s11121-007-0070-9

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121–135. doi:10.1007/s11336-008-9098-4

Griffiths, R. R., Johnson, M. W., Richards, W. A., Richards, B. D., McCann, U., & Jesse, R. (2011). Psilocybin occasioned mystical-type experiences: Immediate and persisting dose-related effects. *Psychopharmacology (Berlin)*, *218*(4), 649–665. doi:10.1007/s00213-011-2358-5

Griffiths, R. R., Richards, W. A., Johnson, M. W., McCann, U. D., & Jesse, R. (2008). Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *Journal of Psychopharmacology*, *22*(6), 621–632. doi:10.1177/0269881108094300

Griffiths, R. R., Richards, W. A., McCann, U. D., & Jesse, R. (2006). Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology (Berlin)*, *187*(3), 268–292. doi:10.1007/s00213-006-0457-5

Grinspoon, L., & Bakalar, J. B. (1979). *Psychedelic drugs reconsidered*. New York, NY: Basic Books.

Grob, C. S., Danforth, A. L., Chopra, G. S., Hagerty, M., McKay, C. R., Halberstadt, A. L., & Greer, G. R. (2011). Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Archives of General Psychiatry*, *68*(1), 71–78. doi:10.1001/archgenpsychiatry.2010.116

Grob, C. S. (1994). Psychiatric research with hallucinogens: What have we learned? In C. Rätsch & J. Baker (Eds.), *Yearbook for ethnomedicine* (1st ed., pp. 91–112). Berlin, Germany: VWB.

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, *19*(4), 369–381. doi:10.1037/1040-3590.19.4.369

Gruzdev, N., & Spivak, D. (2006). An exploratory investigation into the association of neuroticization, cognitive style, and spirituality to reported altered states of consciousness in women experiencing childbirth. *The International Journal of Transpersonal Studies*, *25*, 56–61. Retrieved from http://altstates.net/en/hbi/gruzdev-neuroticizationaasc-2006/pdf

Guadagnoli, E., & Velicer, W. F. (1991). A comparison of pattern-matching indexes. *Multivariate Behavioral Research*, *26*(2), 323–343. doi:10.1207/s15327906mbr2602_7

Gurka, M. J., Edwards, L. J., Muller, K. E., & Kupper, L. L. (2006). Extending the Box–Cox transformation to the linear mixed model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(2), 273–288. doi:10.1111/j.1467-985X.2005.00391.x

Guzmán, G., Allen, J. W., & Gartz, J. (2000). A worldwide geographical distribution of the neurotropic fungi, an analysis and a discusstion. *Annali di Museo Civico di Rovereto*, *14*(1998), 189–280. Retrieved from http://www.museocivico.rovereto.tn.it/UploadDocs/104_art09-guzman%20&%20c.pdf

Habermeyer, E. (1999). *Abnorme Bewusstseinszustände bei Gesunden und bei Patienten mit floriden endogenen Psychosen: Eine empirische Untersuchung [Abnormal states of consciousness in healthy subjects and patients with acute endogenous psychosis: An empirical study]*. (Unpublished doctoral dissertation, Rhenish-Westphalian Technical University, Aachen, Germany).

Halpern, J. H., & Pope, H. G. (2003). Hallucinogen persisting perception disorder: What do we know after 50 years? *Drug and Alcohol Dependence*, *69*(2), 109–119. doi:10.1016/S0376-8716(02)00306-X

Hasler, F., Bourquin, D., Brenneisen, R., Bär, T., & Vollenweider, F. X. (1997). Determination of psilocin and 4-hydroxyindole-3-acetic acid in plasma by HPLC-ECD and pharmacokinetic profiles of oral and intravenous psilocybin in man. *Pharmaceutica Acta Helvetiae*, *72*(3), 175–184. doi:10.1016/S0031-6865(97)00014-9

Hasler, F., Bourquin, D., Brenneisen, R., & Vollenweider, F. X. (2002). Renal excretion profiles of psilocin following oral administration of psilocybin: A controlled study in man. *Journal of Pharmaceutical and Biomedical Analysis*, *30*(2), 331–339. doi:10.1016/S0731-7085(02)00278-9

Hasler, F., Grimberg, U., Benz, M. A., Huber, T., & Vollenweider, F. X. (2004). Acute psychological and physiological effects of psilocybin in healthy humans: A double-blind, placebo-controlled dose-effect study. *Psychopharmacology (Berlin)*, *172*(2), 145–156. doi:10.1007/s00213-003-1640-6

Hasler, F., Quednow, B. B., Treyer, V., Schubiger, P. A., Buck, A., & Vollenweider, F. X. (2009). Role of prefrontal serotonin-2A receptors in self-experience during psilocybin induced altered states. *Neuropsychobiology*, *59*(2), 60–60. Annual Conference of the Swiss Society of Sleep Research, Sleep Medicine and Chronobiology (SSSSC) and the Swiss Society of Biological Psychiatry (SSBP). doi:10.1159/000209314

Hasler, F., Studerus, E., Lindner, K., Ludewig, S., & Vollenweider, F. X. (2009). Investigation of serotonin-1A receptor function in the human psychopharmacology of MDMA. *Journal of Psychopharmacology*, *23*(8), 923–935. doi:10.1177/0269881108094650

Heekeren, K., Daumann, J., Neukirch, A., Stock, C., Kawohl, W., Norra, C., … Gouzoulis-Mayfrank, E. (2008). Mismatch negativity generation in the human $5HT_{2A}$ agonist and NMDA antagonist model of psychosis. *Psychopharmacology (Berlin)*, *199*(1), 77–88. doi:10.1007/s00213-008-1129-4

Heimann, H. (1961). Ausdrucksphänomenologie der Modellpsychosen (Psilocybin). Vergleich mit Selbstschilderungen und psychischem Leistungsabfall. [Expressive phenomenology of model psychoses (psilocybin). Comparison with self-description and psychic deficiency of performance]. *Psychiatria et Neurologia*, *141*, 69–100. doi:10.1159/000131192

Heimann, H. (1962). Zur Behandlung therapieresistenter Neurosen mit Modellpsychosen (Psilocybin) [On the treatment of therapy-resistant neuroses with model psychoses (psilocybin)]. *Schweizer Archiv für Neurologie Neurochirurgie und Psychiatrie*, *89*, 214–220. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=2700

Hemsley, D. R., & Ward, E. S. (1985). Individual differences in reaction to the abuse of LSD. *Personality and Individual Differences*, *6*(4), 515–517. doi:10.1016/0191-8869(85)90148-5

Hermle, L., Fünfgeld, M., Oepen, G., Botsch, H., Borchardt, D., Gouzoulis, E., … Spitzer, M. (1992). Mescaline-induced psychopathological, neuropsychological, and neurometabolic effects in normal subjects: Experimental psychosis as a tool for psychiatric research. *Biological Psychiatry*, *32*(11), 976–991. doi:10.1016/0006-3223(92)90059-9

Hermle, L., Kovar, K. A., Hewer, W., & Ruchsow, M. (2008). Halluzinogen-induzierte psychische Störungen [Hallucinogen-induced psychological disorders]. *Fortschritte der Neurologie – Psychiatrie*, *76*(6), 334–342. doi:10.1055/s-2008-1038191

Hermle, L., Spitzer, M., Borchardt, D., Kovar, K. A., & Gouzoulis, E. (1993). Psychological effects of MDE in normal subjects. Are entactogens a new class of psychoactive agents? *Neuropsychopharmacology*, *8*(2), 171–176. Retrieved from http://www.erowid.org/references/texts/show/6147docid5769

Heymans, M. W., van Buuren, S., Knol, D. L., van Mechelen, W., & de Vet, H. C. W. (2007). Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology*, *7*(33). doi:10.1186/1471-2288-7-33

Hillebrand, J., Olszewski, D., & Sedefov, R. (2006). *EMCDDA thematic papers - Hallucinogenic mushrooms: An emerging trend case study*. European Monitoring Centre for Drugs and Drugs Addiction. Retrieved from http://www.emcdda.europa.eu/html.cfm/index31208EN.html

Hill, R. M., Fischer, R., & Warshay, D. (1969). Effects of excitatory and tranquilizing drugs on visual perception. Spatial distortion thresholds. *Experientia*, *25*(2), 171–172. doi:10.1007/BF01899105

Hobbs, J. (2007). *The medical history of psychedelic drugs*. (Unpublished doctoral dissertation, University of Cambridge, England). Retrieved from http://psychedelic.nfshost.com/history_of_psychedelics.pdf

Hofmann, A. (1968). Psychotomimetic agents. In A. Burger (Ed.), *Chemical constitution and pharmacodynamic actions* (Vol. 2, 5, pp. 169–235). New York, NY: M. Dekker.

Hofmann, A., Heim, R., Brack, A., & Kobel, H. (1958). Psilocybin, ein psychotroper Wirkstoff aus dem mexikanischen Rauschpilz Psilocybe mexicana Heim [Psilocybin, a psychotropic substance from the Mexican mushroom Psilocybe mexicana Heim]. *Experientia*, *14*(3), 107–109. doi:10.1007/BF02159243

Hofmann, A., Heim, R., Brack, A., Kobel, H., Frey, A., Ott, H., … Troxler, F. (1959). Psilocybin und Psilocin, zwei psychotrope Wirkstoffe aus mexikanischen Rauschpilzen [Psilocybin and psilocin, two psychotropic substances from hallucinogenic Mexican fungi]. *Helvetica Chimica Acta*, *42*, 1557–1572. doi:10.1002/hlca.19590420518

Hofmann, A. (2003). Podiumsdiskussion "Was ist Bewußtseinserweiterung?", Heidelberg, Februar 1996; track 12. On *Erinnerungen eines Psychonauten: Von der Entdeckung entheogener Drogen* [CD]. Köln, Germany: supposé.

Hollister, L. E. (1961). Clinical, biochemical and psychologic effects of psilocybin. *Archives Internationales de Pharmacodynamie et de Therapie*, *130*, 42–52. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=2622

Hollister, L. E., & Hartman, A. M. (1962). Mescaline, lysergic acid diethylamide and psilocybin comparison of clinical syndromes, effects on color perception and biochemical measures. *Comprehensive Psychiatry*, *3*, 235–242. doi:10.1016/S0010-440X(62)80024-8

Holstein, D. H., Csomor, P. A., Geyer, M. A., Huber, T., Brugger, N., Studerus, E., & Vollenweider, F. X. (2011). The effects of sertindole on sensory gating, sensorimotor gating, and cognition in healthy volunteers. *Journal of Psychopharmacology*, *25*(12), 1600–1613. doi:10.1177/0269881111415734

Holsten, F. (1976). Flashbacks: A personal follow-up. *European Archives of Psychiatry and Clinical Neuroscience*, *222*(4), 293–304. doi:10.1007/BF00343238

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi:10.1007/BF02289447

Horowitz, M. J. (1969). Flashbacks: Recurrent intrusive images after the use of LSD. *American Journal of Psychiatry*, *126*(4), 565–569. doi:10.1176/appi.ajp.126.4.565

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. doi:10.1002/bimj.200810425

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, *6*(6), 153–160. doi:10.1007/BF02289270

Hübner, C. (2007). EnTrance: Entrance to wider worlds, or mystification of mere relaxation? *Music Therapy Today*, *8*(2), 257–293. Retrieved from http://www.wfmt.info/Musictherapyworld/modules/mmmagazine/issues/20070718101131/20070718101842/MTT8_2_Huebner.pdf

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453. doi:10.1037/1082-989X.3.4.424

Huxley, A. (1959). *The doors of perception and heaven and hell*. Harmondsworth, England: Penguin.

Huysamen, G. K. (2006). Coefficient alpha: Unnecessarily ambiguous; unduly ubiquitous. *SA Journal of Industrial Psychology*, *32*(4), 34–40. Retrieved from http://sajip.co.za/index.php/sajip/article/viewFile/242/239

Hyde, R. W. (1960). Psychological and social determinants of drug action. In G. J. Sarwer-Foner (Ed.), *The dynamics of psychiatric drug therapy* (pp. 297–315). Springfield, IL: Thomas.

Hysek, C. M., Brugger, R., Simmler, L. D., Bruggisser, M., Donzelli, M., Grouzmann, E., ... Liechti, M. E. (2012). Effects of the α2-adrenergic agonist clonidine on the pharmacodynamics and pharmacokinetics of 3,4 methylenedioxymethamphetamine in healthy volunteers. *The Journal of Pharmacology and Experimental Therapeutics*, *340*(2), 286–294. doi:10.1124/jpet.111.188425

Hysek, C. M., Simmler, L. D., Ineichen, M., Grouzmann, E., Hoener, M. C., Brenneisen, R., ... Liechti, M. E. (2011). The norepinephrine transporter inhibitor reboxetine reduces stimulant effects of MDMA ("ecstasy") in humans. *Clinical Pharmacology and Therapeutics*, *90*, 246–255. doi:10.1038/clpt.2011.78

Isbell, H. (1959). Comparison of the reactions induced by psilocybin and LSD-25 in man. *Psychopharmacologia*, *1*, 29–38. doi:10.1007/BF00408109

Iwata, N., & Higuchi, H. R. (2000). Responses of Japanese and American university students to the STAI items that assess the presence or absence of anxiety. *Journal of Personality Assessment*, *74*(1), 48–62. doi:10.1207/S15327752JPA740104

Janke, W. (Ed.). (1983). *Response variability to psychotropic drugs*. New York, NY: Pergamon.

Janke, W., & Debus, G. (1978). *Die Eigenschaftswörterliste EWL. Eine mehrdimensionale Methode zur Beschreibung von Aspekten des Befindens [The adjective word list EWL. A multidimensional method for describing aspects of the actual state]* (1st ed.). Göttingen, Germany: Hogrefe.

Janke, W., & Debus, G. (1986). EWL 60 S. In *Internationale Skalen für Psychiatrie [International scales for psychiatry]* (3rd ed., pp. 43–44). Weinheim, Germany: Beltz.

Johnson, M. W., Andrew Sewell, R., & Griffiths, R. R. (2011). Psilocybin dose-dependently causes delayed, transient headaches in healthy volunteers. *Drug and Alcohol Dependence*. doi:10.1016/j.drugalcdep.2011.10.029

Johnson, M. W., Richards, W. A., & Griffiths, R. R. (2008). Human hallucinogen research: Guidelines for safety. *Journal of Psychopharmacology*, *22*(6), 603–620. doi:10.1177/0269881108093587

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. doi:10.1007/BF02289343

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. doi:10.1007/BF02291366

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL. Revised version.* Retrieved from http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf

Kjellgren, A., Bood, S. A., Axelsson, K., Norlander, T., & Saatcioglu, F. (2007). Wellness through a comprehensive yogic breathing program - A controlled pilot trial. *BMC Complementary and Alternative Medicine*, *7*(43). doi:10.1186/1472-6882-7-43

Kjellgren, A., & Buhrkall, H. (2010). A comparison of the restorative effect of a natural environment with that of a simulated natural environment. *Journal of Environmental Psychology*, *30*(4), 464 –472. doi:10.1016/j.jenvp.2010.01.011

Kjellgren, A., Erdefelt, K., Werngren, L., & Norlander, T. (2011). Does relaxation on a bed of nails (spike mat) induce beneficial effects? A randomized controlled pilot study. *Alternative Medicine Studies*, *1*(e5). doi:10.4081/ams.2011.e5

Kjellgren, A., Sundequist, U., Norlander, T., & Archer, T. (2001). Effects of flotation-REST on muscle tension pain. *Pain Research & Managment*, *6*(4), 181–189.

Kjellgren, A., Sundequist, U., Sundholm, U., Norlander, T., & Archer, T. (2004). Altered consciousness in flotation-REST and chamber-REST: Experience of experimental pain and subjective stress. *Social Behavior and Personality*, *32*(2), 103–115. doi:10.2224/sbp.2004.32.2.103

Kometer, M., Cahn, B. R., Andel, D., Carter, O. L., & Vollenweider, F. X. (2011). The 5-$HT_{2A/1A}$ agonist psilocybin disrupts modal object completion associated with visual hallucinations. *Biological Psychiatry*, *69*(5), 339–406. doi:10.1016/j.biopsych.2010.10.002

Kornetsky, C., & Humphries, O. (1957). Relationships between effects of a number of centrally acting drugs and personality. *AMA Archives of Neurology and Psychiatry*, *77*(3), 325–327. doi:10.1001/archneurpsyc.1957.02330330111017

Korth, B., & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data. *Psychometrika*, *40*(3), 361–372. doi:10.1007/BF02291763

Koukkou, M., & Lehmann, D. (1976). Human EEG spectra before and during cannabis hallucinations. *Biological Psychiatry*, *11*(6), 663–677.

Koukkou, M., & Lehmann, D. (1978). Korrelationen zwischen Cannabis-induzierter Psychopathologie und EEG vor und nach Drogen-Einnahme [Correlations between cannabis-induced psychopathology and EEG before and after drug ingestion]. *Pharmakopsychiatry*, *11*(5), 220–227. doi:10.1055/s-0028-1094581

Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science*, *45*(1), 106–110. Retrieved from http://www.pabst-publishers.de/psychology-science/1-2003/pdf_06.pdf

Kvaal, K., Laake, K., & Engedal, K. (2001). Psychometric properties of the state part of the Spielberger State-Trait Anxiety Inventory (STAI) in geriatric patients. *International Journal of Geriatric Psychiatry*, *16*(10), 980–986. doi:10.1002/gps.458

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363–374. doi:10.2307/2529786

Langlitz, N. (2007). *Neuropsychedelia. The Revival of Hallucinogen Research since the Decade of the Brain.* (Unpublished PhD thesis, University of California, Berkeley). Retrieved from http://www.erowid.org/references/texts/show/7374docid6535

Langlitz, N. (2010). The persistence of the subjective in neuropsychopharmacology: observations of contemporary hallucinogen research. *History of Human Sciences*, *23*(1), 37–57. doi:10.1177/0952695109352413

Langlitz, N. (in press). Delirious brain chemistry and controlled culture. Exploring the contextual mediation of drug effects. In S. Choudhury & J. Slaby (Eds.), *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*. Oxford, UK: Blackwell.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar [The State-Trait Anxiety Inventory]*. Weinheim, Germany: Beltz.

Leary, T. (1964). How to change behavior. In D. Solomon (Ed.), *LSD: The consciousness-expanding drug* (pp. 103–118). New York, NY: Putnam. Retrieved from http://www.scribd.com/doc/9638444/LSD-The-ConsciousnessExpanding-Drug

Leary, T., Litwin, G. H., & Metzner, R. (1963). Reactions to psilocybin administered in a supportive environment. *The Journal of Nervous and Mental Disease*, *137*(6), 561–573. Retrieved from http://www.erowid.org/references/texts/show/2688docid2311

Leuner, H. (1962). *Die experimentelle Psychose [The experimental psychosis]*. Berlin, Germany: Springer.

Leuner, H. (1968). Ist die Verwendung von LSD-25 für die experimentelle Psychiatrie und in der Psychotherapie heute noch vertretbar? [Is the use of LSD-25 in experimental psychiatry and psychotherapy today still justifiable?] *Nervenarzt*, *39*(8), 356–360. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=4702

Leuner, H. (1971). Halluzinogene in der Psychotherapie [Hallucinogens in psychotherapy]. *Pharmacopsychiatry*, *4*, 333–351. doi:10.1055/s-0028-1094326

Leuner, H. (1981). *Halluzinogene – Psychische Grenzzustände in Forschung und Psychotherapie [Hallucinogens – Psychological borderline states in research and psychotherapy]*. Berne, Switzerland: Huber.

Liechti, M. E., & Vollenweider, F. X. (2001). Which neuroreceptors mediate the subjective effects of MDMA in humans? A summary of mechanistic studies. *Human Psychopharmacoly*, *16*(8), 589–598. doi:10.1002/hup.348

Liechti, M. E., Baumann, C., Gamma, A., & Vollenweider, F. X. (2000). Acute psychological effects of 3,4-methylenedioxymethamphetamine (MDMA, "Ecstasy") are attenuated by the serotonin uptake inhibitor citalopram. *Neuropsychopharmacology*, *22*(5), 513–521. doi:10.1016/S0893-133X(99)00148-7

Liechti, M. E., Gamma, A., & Vollenweider, F. X. (2001). Gender differences in the subjective effects of MDMA. *Psychopharmacology (Berlin)*, *154*(2), 161–168. doi:10.1007/s002130000648

Liechti, M. E., Geyer, M. A., Hell, D., & Vollenweider, F. X. (2001). Effects of MDMA (ecstasy) on prepulse inhibition and habituation of startle in humans after pretreatment with citalopram, haloperidol, or ketanserin. *Neuropsychopharmacology*, *24*(3), 240–252. doi:10.1016/S0893-133X(00)00199-8

Liechti, M. E., Saur, M. R., Gamma, A., Hell, D., & Vollenweider, F. X. (2000). Psychological and physiological effects of MDMA ("Ecstasy") after pretreatment with the 5-HT(2) antagonist ketanserin in healthy humans. *Neuropsychopharmacology*, *23*(4), 396–404. doi:10.1016/S0893-133X(00)00126-3

Liechti, M. E., & Vollenweider, F. X. (2000). Acute psychological and physiological effects of MDMA ("ecstasy") after haloperidol pretreatment in healthy humans. *European Neuropsychopharmacology*, *10*(4), 289–295. doi:10.1016/S0924-977X(00)00086-9

Lienert, G. A., & Netter, P. (1996). LSD response in Eysenckian trait types identified by polypredictive CFA. *Personality and Individual Differences*, *21*(6), 845–850. doi:10.1016/S0191-8869(96)00143-2

Linton, H. B., Langs, R. J., & Paul, I. H. (1964). Retrospective alterations of the LSD-25 experience. *The Journal of Nervous and Mental Disease*, *138*(5), 409–423. Retrieved from http://www.erowid.org/references/texts/show/2364docid2145

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151–173. doi:10.1207/S15328007SEM0902_1

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coeffcient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. doi:10.1027/1614-2241.2.2.57

Ludwig, A. M. (1966). Altered states of consciousness. *Archives of General Psychiatry*, *15*(3), 225–234. doi:10.1001/archpsyc.1966.01730150001001

Maclean, K. A., Johnson, M. W., & Griffiths, R. R. (2011). Mystical experiences occasioned by the hallucinogen psilocybin lead to increases in the personality domain of openness. *Journal of Psychopharmacology*, *25*(11), 1453–1461. doi:10.1177/0269881111420188

Malitz, S., Esecover, H., Wilkens, B., & Hoch, P. H. (1960). Some observations on psilocybin, a new hallucinogen, in volunteer subjects. *Comprehensive Psychiatry*, *1*, 8–17. doi:10.1016/S0010-440X(60)80045-4

Malitz, S., Wilkens, B., & Esecover, H. (1962). A comparison of drug-induced hallucinations with those seen in spontaneously occurring psychoses. In C. West (Ed.), *Hallucinations: behaviour, experience and theory* (pp. 50–63). New York, NY: Grune & Stratton.

El-Mallakh, R. S., Halpern, J. H., & Abraham, H. D. (2008). Substance abuse: Hallucinogen- and MDMA-related disorders. In A. Tasman, M. Maj, M. First, J. Kay & J. Lieberman (Eds.), *Psychiatry* (3rd ed., pp. 1100–1126). New York, NY: Wiley. doi:10.1002/9780470515167.ch60

Mangini, M. (1998). Treatment of alcoholism using psychedelic drugs: A review of the program of research. *Journal of Psychoactive Drugs*, *30*(4), 381–418. Retrieved from http://www.erowid.org/references/texts/show/6982docid6372

Marlatt, G. A., & Rohsenow, D. J. (1980). Cognitive processes in alcohol use: Expectancy and the balanced placebo design. In N. K. Mello (Ed.), *Advances in substance abuse: behavioural and biological research.* (pp. 159–199). Greenwich, CT: JAI Press.

Marsh, H. W., & Hau, K. T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, *32*(1), 151–170. Applications of Latent Variable Modeling in Educational Psychology Research. doi:10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491. doi:10.1037/a0019227

Marsh, H. W., & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, *64*(4), 364–390. Retrieved from http://www.jstor.org/stable/20152499

Marsh, H. W., Muthén, B. O., Asparouhov, T., Ludtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*(3), 439–476. doi:10.1080/10705510903008220

Masters, R. E. L., & Houston, J. (1967). *The varieties of psychedelic experience*. London: Anthony Blond.

Mavromatis, A. (1987). *Hypnagogia: The unique state of consciousness between wakefulness and sleep* (1st ed.). London: Routledge and Kegan Paul.

McCrae, R. R., Zonderman, A. B., Bond, M. H., Costa, P. T., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566. doi:10.1037/0022-3514.70.3.552

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McGlothlin, W. H., & Arnold, D. O. (1971). LSD revisited. A ten-year follow-up of medical LSD use. *Archives of General Psychiatry*, *24*(1), 35–49. doi:10.1001/archpsyc.1971.01750070037005

McGlothlin, W. H., Cohen, S., & McGlothlin, M. S. (1967). Long lasting effects of LSD on normals. *Archives of General Psychiatry*, *17*(5), 521–532. doi:10.1001/archpsyc.1967.01730290009002

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). Profile of mood states (POMS). *San Diego: Educational and Industrial Testing Service*.

Mendelson, J. E., Coyle, J. R., Lopez, J. C., Baggott, M. J., Flower, K., Everhart, E. T., … Cohen, B. M. (2011). Lack of effect of sublingual salvinorin A, a naturally occurring kappa opioid, in humans: A placebo-controlled trial. *Psychopharmacology (Berlin)*, *214*(4), 933–939. doi:10.1007/s00213-010-2103-5

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi:10.1007/BF02294825

Metrik, J., Rohsenow, D. J., Monti, P. M., McGeary, J., Cook, T. A. R., de Wit, H., … Kahler, C. W. (2009). Effectiveness of a marijuana expectancy manipulation: Piloting the balanced-placebo design for marijuana. *Experimental and Clinical Psychopharmacology*, *17*(4), 217–225. doi:10.1037/a0016502

Metzner, R., Litwin, G. H., & Weil, G. M. (1965). The relation of expectation and mood to psilocybin reactions: A questionnaire study. *Psychedelic Review*, *5*, 3–39. Retrieved from http://www.maps.org/psychedelicreview/v1n5/01503met.pdf

Millsap, R. E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In *Factor analysis at 100: historical developments and future directions* (pp. 131–152). Mahwah, NJ: Lawrence Erlbaum.

Moreau de Tours, J. (1845). *Du hachisch et de l'aliénation mentale - Etudes psychologiques*. Paris, France: Librairie de Fortin.

Moreno, F. A., Wiegand, C. B., Taitano, E. K., & Delgado, P. L. (2006). Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *Journal of Clinical Psychiatry*, *67*(11), 1735–1740. Retrieved from http://www.psychiatrist.com/abstracts/200611/110610.htm

Moskowitz, D. (1971). Use of haloperidol to reduce LSD flashbacks. *Military Medicine*, *136*(9), 754–756. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=5257

Musso, F., Brinkmeyer, J., Ecker, D., London, M. K., Thieme, G., Warbrick, T., … Winterer, G. (2011). Ketamine effects on brain function - Simultaneous fMRI/EEG during a visual oddball task. *Neuroimage*, *58*(2), 508–525. doi:10.1016/j.neuroimage.2011.06.045

Muthén, B. O. (1989). Dichotomous factor analysis of symptom data. *Sociological Methods Research*, *18*(1), 19–65. doi:10.1177/0049124189018001002

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. doi:10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide. Fifth edition.* Los Angeles, CA: Muthén & Muthén.

Nesselroade, J., & Baltes, P. B. (1970). On a dilemma of comparative factor analysis: A study of factor matching based on random data. *Educational and Psychological Measurement*, *30*(4), 935–948. doi:10.1177/001316447003000413

Nichols, D. E. (2004). Hallucinogens. *Pharmacology & Therapeutics*, *101*(2), 131–181. doi:10.1016/j.pharmthera.2003.11.002

Nichols, D. E., & Chemel, B. R. (2006). The neuropharmacology of religious experience: Hallucinogens and the experience of the divine. In *Where god and science meet. How brain and evolutionary studies alter our understanding of religion* (1, pp. 1–33). Westport, CT: Prager.

Nieto, D. (1962). Psicosis experimentales con psilocybina [Experimental psychosis with psilocybin]. *Neurologia Neuocirugia-Psiquiatria*, *3*, 140–146. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=2709

Northoff, G., Richter, A., Bermpohl, F., Grimm, S., Martin, E., Marcar, V. L., … Boeker, H. (2005). NMDA hypofunction in the posterior cingulate as a model for schizophrenia: An exploratory ketamine administration study in fMRI. *Schizophrenia Research*, *72*(2-3), 235–248. doi:10.1016/j.schres.2004.04.009

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Nutt, D. J., King, L. A., Saulsbury, W., & Blakemore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse. *Lancet*, *369*(9566), 1047–1053. doi:10.1016/S0140-6736(07)60464-4

O'Brian, C. P. (2005). Drug addiction and drug abuse. In L. Brunton, J. Lazo & K. Parker (Eds.), *Goodman and gilman's the pharmacological basis of therapeutics* (11th ed., pp. 607–628). New York, NY: McGraw-Hill.

Ott, U., Reuter, M., Hennig, J., & Vaitl, D. (2005). Evidence for a common biological basis of the Absorption trait, hallucinogen effects, and positive symptoms: Epistasis between 5-HT2a and COMT polymorphisms. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, *137B*(1), 29–32. doi:10.1002/ajmg.b.30197

Pahnke, W. N. (1963). *Drugs and mysticism: An analysis of the relationship between psychedelic drugs and the mystical consciousness.* (Doctoral dissertation, Harvard University).

Pahnke, W. N. (1969). Psychedelic drugs and mystical experience. *International Psychiatry Clinics*, *5*(4), 149–162.

Passie, T. (1995). Psilocybin in der modernen Psychotherapie [Psilocybin in modern psychotherapy]. *Curare*, *18*, 131–152.

Passie, T., Karst, M., Borsutzky, M., Wiese, B., Emrich, H. M., & Schneider, U. (2003). Effects of different subanaesthetic doses of (S)-ketamine on psychopathology and binocular depth inversion in man. *Journal of Psychopharmacology*, *17*(1), 51–56. doi:10.1177/0269881103017001698

Passie, T., Seifert, J., Schneider, U., & Emrich, H. M. (2002). The pharmacology of psilocybin. *Addiction Biology*, *7*(4), 357–364. doi:10.1080/1355621021000005937

Paunonen, S. V. (1997). On chance and factor congruence following orthogonal Procrustes rotation. *Educational and Psychological Measurement*, *57*(1), 33–59. doi:10.1177/0013164497057001003

Pekala, R. J., Steinberg, J., & Kumar, V. K. (1986). Measurement of phenomenological experience: Phenomenology of Consciousness Inventory. *Perceptual and motor skills*, *63*(2), 983–989. doi:10.2466/pms.1986.63.2.983

Pekala, R. J., Wenger, C. F., & Levine, R. L. (1985). Individual differences in phenomenological experience: States of consciousness as a function of absorption. *Journal of Personality and Social Psychology*, *48*(1), 125–132. doi:10.1037/0022-3514.48.1.125

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and computing. New York, NY: Springer. doi:10.1007/b98882

Pinheiro, J. C., Bates, D. M., DebRoy, S., Sarkar, D., & the R Core team. (2008). nlme: Linear and nonlinear mixed effects models (Version 3.1-90) [Computer software]. Retrieved from http://CRAN.R-project.org/package=nlme

Pinneau, S. R., & Newhouse, A. (1964). Measures of invariance and comparability in factor-analysis for fixed variables. *Psychometrika*, *29*(3), 271–281. doi:10.1007/BF02289724

Polito, V., Langdon, R., & Brown, J. (2010). The experience of altered states of consciousness in shamanic ritual: The role of pre-existing beliefs and affective factors. *Conscious and Cognition*, *19*(4), 918–925. doi:10.1016/j.concog.2010.05.013

Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Quednow, B. B., Kometer, M., Geyer, M. A., & Vollenweider, F. X. (2012). Psilocybin-induced deficits in automatic and controlled inhibition are attenuated by ketanserin in healthy human volunteers. *Neuropsychopharmacology*, *37*, 630–640. doi:10.1038/npp.2011.228

Quednow, B. B., Treyer, V., Hasler, F., Dörig, N., Wyss, M. T., Burger, C., … Vollenweider, F. X. (2011). Assessment of serotonin release capacity in the human brain using dexfenfluramine challenge and [(18)F]altanserin positron emission tomography. *Neuroimage*. doi:10.1016/j.neuroimage.2011.09.045

Raiche, G., & Magis, D. (2009). nFactors: Parallel analysis and non graphical solutions to the Cattell scree test (Version 2.3.1) [Computer software]. Retrieved from http://CRAN.R-project.org/package=nFactors.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical & Statistical Psychology*, *54*(Pt 2), 315–323. doi:10.1348/000711001159582

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development*, *42*(3), 223–232. doi:10.1177/0748175609344096

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, *11*(4), 621–637. doi:10.1207/s15328007sem1104_7

R Development Core Team. (2008). R: A language and environment for statistical computing (Version 2.8.1) [Computer software]. Retrieved from http://www.R-project.org.

R Development Core Team. (2009). R: A language and environment for statistical computing (Version 2.10.1) [Computer software]. Retrieved from http://www.R-project.org.

R Development Core Team. (2011). *R: a language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from R Foundation for Statistical Computing: http://www.R-project.org/

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287–297. doi:10.1037/1040-3590.12.3.287

Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale construction. *Behavior Research Methods and Instrumentation*, *10*(5), 739–742. doi:10.3758/BF03205389

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74. doi:10.1207/s15327906mbr1401_4

Revelle, W. (2009). psych: Procedures for psychological, psychometric, and personality research (Version 1.0-85) [Computer software]. Retrieved from http://CRAN.R-project.org/package=psych.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*, 403–414. doi:10.1207/s15327906mbr1404_2

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the GLB: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154. doi:10.1007/s11336-008-9102-z

Riba, J., Rodríguez-Fornells, A., & Barbanoj, M. J. (2002). Effects of ayahuasca on sensory and sensorimotor gating in humans as measured by P50 suppression and prepulse inhibition of the startle reflex, respectively. *Psychopharmacology (Berlin)*, *165*(1), 18–28. doi:10.1007/s00213-002-1237-5

Riba, J., Rodríguez-Fornells, A., Strassman, R. J., & Barbanoj, M. J. (2001). Psychometric assessment of the hallucinogen rating scale. *Drug and Alcohol Dependence*, *62*(3), 215–223. doi:10.1016/S0376-8716(00)00175-7

Rinkel, M. A., DiMascio, A., Robey, A., & Atwell, C. (1961). Personality patterns and reaction to psilocybine. In P. Bradley (Ed.), *Neuro-psychopharmacology vol. 2* (pp. 56–85). Amsterdam, The Netherlands: Elsevier.

Ritz, T., & Dahme, B. (1995). Die Absorption-Skala: Konzeptuelle Aspekte, psychometrische Kennwerte und Dimensionalität einer deutschsprachigen Adaptation [The Absorption Scale: Conceptual aspects, psychometric properties, and dimensionality of a German adaptation]. *Diagnostica*, *41*, 53–61.

Roche, S., & McConkey, K. (1990). Absorption: Nature, assessment, and correlates. *Journal of Personality and Social Psychology*, *59*(1), 91–101. doi:10.1037/0022-3514.59.1.91

Rolland, R. (1967). *Un beau visage à tous sens: avec: choix de lettres de romain rolland (1866–1944)*. Paris, France: Editions Albin Michel.

Rosseel, Y., Oberski, D., & Byrnes, J. (2011). lavaan: Latent Variable Analysis (Version 0.4-9) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lavaan.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*(4), 305–335. doi:10.1016/S0167-8116(02)00097-6

Royston, P., & Sauerbrei, W. (2008). *Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for continuous variables*. Wiley series in probability and statistics. Chichester, England: Wiley.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Rümmele, W., & Gnirss, F. (1961). Untersuchungen mit Psilocybin, einer psychotropen Substanz aus Psilocybe Mexicana [Studies on psilocybin, a hallucinogenic drug from the Mexican mushroom Psilocybe Mexicana]. *Schweizer Archiv für Neurologie Neurochirurgie und Psychiatrie*, *87*, 365–385. Retrieved from http://www.erowid.org/references/refs_view.php?A=ShowDoc1&ID=2624

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*(1), 3–30. doi:10.1037/0033-2909.125.1.3

Sanders-Bush, E., Burris, K. D., & Knoth, K. (1988). Lysergic acid diethylamide and 2,5-dimethoxy-4-methylamphetamine are partial agonists at serotonin receptors linked to phosphoinositide hydrolysis. *Journal of Pharmacology and Experimental Therapeutics*, *246*(3), 924–928. Retrieved from http://jpet.aspetjournals.org/content/246/3/924.abstract

Sandison, R. A. (1954). Psychological aspects of the LSD treatment of the neuroses. *Journal of Mental Science*, *100*(419), 508–515. doi:10.1192/bjp.100.419.508

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. doi:10.1007/BF02296192

Sauerbrei, W., Holländer, N., & Buchholz, A. (2008). Investigation about a screening step in model selection. *Statistics and Computing*, *18*, 195–208. 10.1007/s11222-007-9048-5. doi:10.1007/s11222-007-9048-5

Schmidt, A., Csomor, P. A., Bachmann, R., Kometer, M., & Vollenweider, F. X. (2010). Emotional face processing in the 5-HT_{2A} model of altered states of consciousness using event-related potentials. *European Neuropsychopharmacology*, *20*(Supplement 3), S231–S231. Papers of the 23rd ECNP Congress. doi:10.1016/S0924-977X(10)70280-7

Schmidt, A., Bachmann, R., Kometer, M., Csomor, P. A., Stephan, K. E., Seifritz, E., & Vollenweider, F. X. (2011). Mismatch negativity encoding of prediction errors predicts S-ketamine-induced cognitive impairments. *Neuropsychopharmacology*. doi:10.1038/npp.2011.261

Schmitz, N., Hartkamp, N., Kiuse, J., Franke, G. H., Reister, G., & Tress, W. (2000). The Symptom Check-List-90-R (SCL-90-R): A German validation study. *Quality of Life Research*, *9*(2), 185–193. doi:10.1023/A:1008931926181

Schmitz, P. (2004). On the alternative five-factor model: Structure and correlates. In R. M. Stelmack (Ed.), *On the psychobiology of personality: Essays in honor of Marvin Zuckerman* (pp. 65–87). Sand Diego, CA: Elsevier.

Schmukle, S., & Egloff, B. (2009). Exploring bipolarity of affect ratings by using polychoric correlations. *Cognition and Emotion*, *23*(2), 272–295. doi:10.1080/02699930801987330

Schneiter, B. (1991). *Akustische-halluzinatorische Phänomene bei Gesunden: Eine experimentelle Untersuchung mit sensorischer Deprivation, Stickoxydul und N,N-Dimethyl-*

tryptamin [Acoustic-hallucinatory phenomena in healthy subjects: An experimental study with sensory deprivation, nitrous oxide, and N,N-dimethyltryptamine]. (Unpublished master's thesis, University of Zurich, Switzerland).

Segura, S. L., & González-Romá, V. (2003). How do respondents construe ambiguous response formats of affect items? *Journal of Personality and Social Psychology*, *85*, 956–968. doi:10.1037/0022-3514.85.5.956

Sessa, B. (2008). Are psychedelic drug treatments seeing a comeback in psychiatry? *Progress in Neurology and Psychiatry*, *12*(8), 5–10. doi:10.1002/pnp.94

Sewell, R. A., Halpern, J. H., & Pope, H. G. (2006). Response of cluster headache to psilocybin and LSD. *Neurology*, *66*(12), 1920–1922. doi:10.1212/01.wnl.0000219761.05466.43

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. doi:10.1007/s11336-008-9101-0

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *State-Trait Anxiety Inventory for adults*. Palo Alto, CA: Consulting Psychologists.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *STAI-Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists.

Spitzer, M., Franke, B., Walter, H., Buechler, J., Wunderlich, A. P., Schwab, M., … Grön, G. (2001). Enantio-selective cognitive and brain activation effects of N-ethyl-3,4-methylenedioxyamphetamine in humans. *Neuropharmacology*, *41*(2), 263–271. doi:10.1016/S0028-3908(01)00060-0

Spitzer, M., Thimm, M., Hermle, L., Holzmann, P., Kovar, K. A., Heimann, H., … Schneider, F. (1996). Increased activation of indirect semantic associations under psilocybin. *Biological Psychiatry*, *39*(12), 1055–1057. doi:10.1016/0006-3223(95)00418-1

Sprenger, T., Valet, M., Woltmann, R., Zimmer, C., Freynhagen, R., Kochs, E. F., … Wagner, K. J. (2006). Imaging pain modulation by subanesthetic s-(+)-ketamine. *Anesthesia & Analgesia*, *103*(3), 729–737. doi:10.1213/01.ane.0000231635.14872.40

Stace, W. T. (1961). *Mysticism and philosophy*. London: Macmillan.

Stafford, P. G., & Golightly, B. H. (1967). *LSD: The problem-solving psychedelic*. New York, NY: Award Books.

Stamets, P. (1996). *Psilocybin mushrooms of the world an identification guide*. Berkeley, CA: Ten Speed.

Steyerberg, E. W. (2009). *Clinical prediction models: A practical approach to development, validation, and updating*. New York, NY: Springer. doi:10.1007/978-0-387-77244-8

Strassman, R. J. (1984). Adverse reactions to psychedelic drugs. A review of the literature. *Journal of Nervous and Mental Disease*, *172*(10), 577–595. doi:10.1097/00005053-198410000-00001

Strassman, R. J., Qualls, C. R., Uhlenhuth, E. H., & Kellner, R. (1994). Dose-response study of N,N-dimethyltryptamine in humans. II. Subjective effects and preliminary results of a new rating scale. *Archives of General Psychiatry*, *51*(2), 98–108. doi:10.1001/archpsyc.1994.03950020022002

Studerus, E., Gamma, A., Kometer, M., & Vollenweider, F. X. (2012). Prediction of psilocybin response in healthy volunteers. *PLoS ONE*, *7*(2), e30800. doi:10.1371/journal.pone.0030800

Studerus, E., Gamma, A., & Vollenweider, F. X. (2010). Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS ONE*, *5*(8), e12412. doi:10.1371/journal.pone.0012412

Studerus, E., Kometer, M., Hasler, F., & Vollenweider, F. X. (2011). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: A pooled analysis of experimental studies. *Journal of Psychopharmacology*, *25*(11), 1434–1452. doi:10.1177/0269881110382466

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*(3), 268–277. doi:10.1037/h0036681

Thompson, B. (2003). *Score reliability: contemporary thinking on reliability issues*. Thousand Oaks, CA: SAGE.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies*. (Personel Research Section Report No. 984). Washington, DC: Department of the Army.

Umbricht, D., Koller, R., Vollenweider, F. X., & Schmid, L. (2002). Mismatch negativity predicts psychotic experiences induced by NMDA receptor antagonist in healthy volunteers. *Biological Psychiatry*, *51*(5), 400–406. doi:10.1016/S0006-3223(01)01242-2

Umbricht, D., Vollenweider, F. X., Schmid, L., Grübel, C., Skrabo, A., Huber, T., & Koller, R. (2003). Effects of the 5-HT_{2A} agonist psilocybin on mismatch negativity generation and AX-continuous performance task: Implications for the neuropharmacology of cognitive deficits in schizophrenia. *Neuropsychopharmacology*, *28*(1), 170–181. doi:10.1038/sj.npp.1300005

Ungerleider, J. T., Fisher, D. D., Fuller, M., & Caldwell, A. (1968). The "bad trip"– the etiology of the adverse LSD reaction. *American Journal of Psychiatry*, *124*(11), 1483–1490. doi:10.1176/appi.ajp.124.11.1483

Vaitl, D., Birbaumer, N., Gruzelier, J., Jamieson, G. A., Kotchoubey, B., Kübler, A., … Weiss, T. (2005). Psychobiology of altered states of consciousness. *Psychological Bulletin*, *131*(1), 98–127. doi:10.1037/0033-2909.131.1.98

van Amsterdam, J., Opperhuizen, A., & van den Brink, W. (2011). Harm potential of magic mushroom use: A review. *Regulatory Toxicology and Pharmacology*, *59*(3), 423–429. doi:10.1016/j.yrtph.2011.01.006

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–68. Retrieved from http://www.jstatsoft.org/v45/i03

van der Zwaard, R., & Polak, M. A. (2001). Pseudohallucinations: a pseudoconcept? A review of the validity of the concept, related to associate symptomatology. *Comprehensive Psychiatry*, *42*(1), 42–50. doi:10.1053/comp.2001.19752

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. doi:10.1007/BF02293557

Velicer, W. F., Eaton, A. C., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). New York, NY: Kluwer.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231–251. doi:10.1037/1082-989X.3.2.231

Vigneau, F., & Cormier, S. (2008). The factor structure of the State-Trait Anxiety Inventory: An alternative view. *Journal of Personality Assessment*, *90*(3), 280–285. doi:10.1080/00223890701885027

Vogel-Sprott, M., & Fillmore, M. T. (1999). Expectancy and behavioural effects of socially used drugs. In E. I. Kirsch (Ed.), *How expectancies shape experience* (pp. 215–232). Washington, DC: American Psychological Association.

Vollenweider, F. X., Csomor, P. A., Knappe, B., Geyer, M. A., & Quednow, B. B. (2007). The effects of the preferential 5-HT_{2A} agonist psilocybin on prepulse inhibition of startle in healthy human volunteers depend on interstimulus interval. *Neuropsychopharmacology*, *32*(9), 1876–1887. doi:10.1038/sj.npp.1301324

Vollenweider, F. X., & Kometer, M. (2010). The neurobiology of psychedelic drugs: Implications for the treatment of mood disorders. *Nature Review Neuroscience*, *11*(9), 642–651. doi:10.1038/nrn2884

Vollenweider, F. X., Liechti, M. E., & Paulus, M. P. (2005). MDMA affects both error-rate dependent and independent aspects of decision-making in a two-choice prediction task. *Journal of Psychopharmacology*, *19*(4), 366–374. doi:10.1177/0269881105053287

Vollenweider, F. X. (1992). The use of psychotomimetics in schizophrenia research with special emphasis on the PCP/ketamine model psychosis. *SUCHT*, *38*(6), 398–409.

Vollenweider, F. X. (2001). Brain mechanisms of hallucinogens and entactogens. *Dialogues in Clinical Neuroscience*, *3*, 265–279. Retrieved from http://www.dialogues-cns.org/brochures/11/pdf/11.pdf

Vollenweider, F. X., Gamma, A., Liechti, M. E., & Huber, T. (1998). Psychological and cardiovascular effects and short-term sequelae of MDMA ("ecstasy") in MDMA-naïve healthy volunteers. *Neuropsychopharmacology*, *19*(4), 241–251. doi:10.1016/S0893-133X(98)00013-X

Vollenweider, F. X., & Geyer, M. A. (2001). A systems model of altered consciousness: Integrating natural and drug-induced psychoses. *Brain Research Bulletin*, *56*(5), 495–507. doi:10.1016/S0361-9230(01)00646-3

Vollenweider, F. X., Hasler, F., & Kometer, M. (2008). Effects of serotonergic hallucinogens on perception and patterns of altanserin displacement in humans. *Fundamental & Clinical Pharmacology*, *22*(Supplement S2), 113–113. Abstracts of the EPHAR 2008 Congress, Manchester, UK, 13-17 July 2008. doi:10.1111/j.1472-8206.2008.00600.x

Vollenweider, F. X., Leenders, K. L., Scharfetter, C., Antonini, A., Maguire, P., Missimer, J., & Angst, J. (1997). Metabolic hyperfrontality and psychopathology in the ketamine model of psychosis using positron emission tomography (PET) and [^{18}F]fluorodeoxyglucose (FDG). *European Neuropsychopharmacology*, *7*(1), 9–24. doi:10.1016/S0924-977X(96)00039-9

Vollenweider, F. X., Leenders, K. L., Scharfetter, C., Maguire, P., Stadelmann, O., & Angst, J. (1997). Positron emission tomography and fluorodeoxyglucose studies of metabolic hyperfrontality and psychopathology in the psilocybin model of psychosis. *Neuropsychopharmacology*, *16*(5), 357–372. doi:10.1016/S0893-133X(96)00246-1

Vollenweider, F. X., Maguire, R. P., Leenders, K. L., Mathys, K., & Angst, J. (1998). Effects of high amphetamine dose on mood and cerebral glucose metabolism in normal volunteers using positron emission tomography (PET). *Psychiatry Research: Neuroimaging*, *83*(3), 149–162. doi:10.1016/S0925-4927(98)00033-X

Vollenweider, F. X., Remensberger, S., Hell, D., & Geyer, M. A. (1999). Opposite effects of 3,4-methylenedioxymethamphetamine (MDMA) on sensorimotor gating in rats versus healthy humans. *Psychopharmacology (Berlin)*, *143*(4), 365–372. doi:10.1007/s002130050960

Vollenweider, F. X., Vollenweider-Scherpenhuyzen, M. F., Bäbler, A., Vogel, H., & Hell, D. (1998). Psilocybin induces schizophrenia-like psychosis in humans via a serotonin-2 agonist action. *Neuroreport*, *9*(17), 3897–3902. Retrieved from http://journals.lww.com/neuroreport/Abstract/1998/12010/Psilocybin_induces_schizophrenia_like_psychosis_in.24.aspx

Vollenweider, F. X., Vontobel, P., Hell, D., & Leenders, K. L. (1999). 5-HT modulation of dopamine release in basal ganglia in psilocybin-induced psychosis in man: A PET study with [^{11}C]raclopride. *Neuropsychopharmacology*, *20*(5), 424–433. doi:10.1016/S0893-133X(98)00108-0

Vollenweider, F. X., Vontobel, P., Oye, I., Hell, D., & Leenders, K. L. (2000). Effects of (S)-ketamine on striatal dopamine: a [^{11}C]raclopride PET study of a model psychosis in humans. *Journal of Psychiatric Research*, *34*(1), 35–43. doi:10.1016/S0022-3956(99)00031-X

von Zerssen, D. (1971). Die Beschwerden-Liste als Test [The List of Complaints as a test]. *Therapiewoche*, *25*, 1908–1920.

Wackermann, J., Wittmann, M., Hasler, F., & Vollenweider, F. X. (2008). Effects of varied doses of psilocybin on time interval reproduction in human subjects. *Neuroscience Letters*, *435*(1), 51–55. doi:10.1016/j.neulet.2008.02.006

Walach, H., & Käseberg, E. (1998). Mind machines: A controlled study on the effects of electromagnetic and optic-acoustic stimulation on general well-being, electrodermal activity, and exceptional psychological experiences. *Behavioral Medicine*, *24*(3), 107–114. doi:10.1080/08964289809596388

Walsh, R. N., & Grob, C. S. (Eds.). (2005). *Higher wisdom: Eminent elders explore the continuing impact of psychedelics*. Albany, NY: SUNY.

Wark, C., & Galliher, J. F. (2010). Timothy Leary, Richard Alpert (Ram Dass) and the changing definition of psilocybin. *International Journal of Drug Policy*, *21*(3), 234–239. doi:10.1016/j.drugpo.2009.08.004

Wasson, R. G. (1958). Les premières sources. In R. Heim, R. Wasson & A. Hofmann (Eds.), *Les champignons hallucinogenes du Mexique [The hallucinogenic mushrooms of Mexico]* (pp. 15–44). Paris, France: Editions du Museum National d'Histoire Naturelle.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

Weber, C. S., Thayer, J. F., Rudat, M., Perschel, F. H., Buchholz, K., & Deter, H. C. (2007). Emotional irritation before mental stress is associated with enhanced peripheral norepinephrine. *Scandinavian Journal of Psychology*, *48*(6), 459–466. doi:10.1111/j.1467-9450.2007.00612.x

Weil, A. T. (1972). Altered states of consciousness. *New Scientist*, 696–698.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer. doi:10.1007/978-0-387-98141-3

Wiswede, D., Münte, T. F., Krämer, U. M., & Rüsseler, J. (2009). Embodied emotion modulates neural signature of performance monitoring. *PLoS ONE*, *4*(6), e5754. doi:10.1371/journal.pone.0005754

Wittmann, M., Carter, O., Hasler, F., Cahn, B. R., Grimberg, U., Spring, P., … Vollenweider, F. X. (2007). Effects of psilocybin on time perception and temporal control of behaviour in humans. *Journal of Psychopharmacology*, *21*(1), 50–64. doi:10.1177/0269881106065859

Wolbach, A. B., Miner, E. J., & Isbel, H. (1962). Comparison of psilocin with psilocybin, mescaline and LSD-25. *Psychopharmacologia*, *3*, 219–223. doi:10.1007/BF00412109

Woods, C. M. (2002). Factor analysis of scales composed of binary items: Illustration with the Maudsley Obsessional Compulsive Inventory. *Journal of Psychopathology and Behavioral Assessment*, *24*(4), 215–223. doi:10.1023/A:1020770831134

Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, *44*(1), 1–27. doi:10.1080/00273170802620121

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the Schedule for Nonadaptive and Adaptive Personality. *Journal of Psychopathology and Behavioral Assessment*, *31*(4), 320–330. doi:10.1007/s10862-008-9118-9

Würmle, O. (1977). Eine experimentelle Untersuchung über das Auftreten abnormer psychischer Zustände bei autogenem Training im Vergleich mit sensorischer Deprivation [Experimental examination on the development of altered states of consciousness by autogenic training in comparison to sensory deprivation]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *25*(3), 264–272.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology*, *30*, 165–200. doi:10.1111/0081-1750.00078

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. (Doctoral dissertation, University of California, Los Angeles). Retrieved from http://www.statmodel.com/download/Yudissertation.pdf

Zhou, X., & Reiter, J. P. (2010). A note on Bayesian inference after multiple imputation. *The American Statistician*, *64*(2), 159–163. doi:10.1198/tast.2010.09109

Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating ω_h for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, *31*(2), 135–157. doi:10.1177/0146621606291558

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. doi:10.1007/s11336-003-0974-7

Zuckerman, M. (2002). Zuckerman-Kuhlman personality questionnaire (ZKPQ): An alternative five-factorial model. In B. De Raad & M. Perugini (Eds.), *Big five assessment* (pp. 377–396). Seattle, WA: Hogrefe & Huber.

Zumbo, B. D., Sireci, S. G., & Hambleton, R. K. (2003). Re-visiting exploratory methods for construct comparability: Is there something to be gained from the ways of

old? Paper presented in the Symposium *Construct Comparability Research: Methodological Issues and Results* organized by Kadriye Ercikan, for the National Council on Measurement in Education (NCME) meetings, April 2003, Chicago, Illinois. Retrieved from http://educ.ubc.ca/faculty/zumbo/aera/papers/Zumbo_Sireci_Hambleton_April_19.pdf

CURRICULUM VITAE

PERSONAL DATA

Name:	Erich Studerus
Date of birth:	20.08.1977
Place of birth:	Münsterlingen TG
Nationality:	Swiss
Address:	Oberwilerstr. 157, CH-4054 Basel
Telephone:	+41 (0)44 241 31 54
Email:	erich.studerus@upkbs.ch

EDUCATION

12/2004 Master of Arts, University of Zurich

- *Major:* Psychology
- *Minor 1:* Psychopathology of Adulthood
- *Minor 2:* Philosophy
- *Specialization:* Clinical Psychology and Psychotherapy (Prof. Dr. Ehlert)
- *Master thesis:* Der $5HT_{1A}$-Rezeptor im psychopharmakologischen Wirkungsprofil von MDMA (Ecstasy) [The $5HT_{1A}$ receptor in the psychopharmacological action profile of MDMA (ecstasy)]

06/1997 Matura Type B, Gymnasium St. Antonius, Appenzell

WORK EXPERIENCE

10/2010 – present	*University Psychiatric Outpatient Department, Psychiatric University Clinics Basel:* Research psychologist
01/2007 – 10/2010	*Department of Neuropsychopharmacology and Brain Imaging, University Hospital of Psychiatry Zurich:* PhD student
03/2010 – 05/2010	*Serranetga GmbH, Zurich:* Freelance programmer

04/2005 – 07/2009	*Department of Research, Psychiatric Services of Aargau:* Research assistant (20% time)
03/2000 – 01/2007	*Sleep Laboratory Fluntern, Zurich:* Somnography assistant (20-40% time)
04/1999 – 05/2006	*Hotel Dolder Waldhaus, Zurich:* Night receptionist (20-40% time)

PEER-REVIEWED ARTICLES

Studerus, E., Gamma, A., Kometer, M., & Vollenweider, F. X. (2012). Prediction of psilocybin response in healthy volunteers. *PLoS ONE*, *7*(2), e30800. doi:10.1371/journal.pone.0030800.

Studerus, E., Kometer, M., Hasler, F., & Vollenweider, F. X. (2011). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: A pooled analysis of experimental studies. *Journal of Psychopharmacology*, *25*(11), 1434–1452. doi:10.1177/0269881110382466.

Studerus, E., Gamma, A., & Vollenweider, F. X. (2010). Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS ONE*, *5*(8), e12412. doi:10.1371/journal.pone.0012412.

Hasler, F., Studerus, E., Lindner, K., Ludewig, S., & Vollenweider, F. X. (2009). Investigation of serotonin-1A receptor function in the human psychopharmacology of MDMA. *Journal of Psychopharmacology*, *23*(8), 923–935. doi:10.1177/0269881108094650.

Holstein, D. H., Csomor, P. A., Geyer, M. A., Huber, T., Brugger, N., Studerus, E., & Vollenweider, F. X. (2011). The effects of sertindole on sensory gating, sensorimotor gating, and cognition in healthy volunteers. *Journal of Psychopharmacology*, *25*(12), 1600–1613. doi:10.1177/0269881111415734.

Csomor, P. A., Yee, B. K., Feldon, J., Theodoridou, A., Studerus, E., & Vollenweider, F. X. (2009). Impaired prepulse inhibition and prepulse-elicited reactivity but intact reflex circuit excitability in unmedicated schizophrenia patients: A comparison with healthy subjects and medicated schizophrenia patients. *Schizophrenia Bulletin*, *35*(1), 244–255. doi:10.1093/schbul/sbm146.

COLOPHON

This thesis was typeset with LaTeX 2ε using Hermann Zapf's *Palatino* and *Euler* type faces (Type 1 PostScript fonts *URW Palladio L* and *FPL* were used). The listings are typeset in *Bera Mono*, originally developed by Bitstream, Inc. as "Bitstream Vera". (Type 1 PostScript fonts were made available by Malte Rosenau and Ulrich Dirr.)

The typographic style was inspired by Bringhurst's genius as presented in *The Elements of Typographic Style* (Bringhurst, 2002). It is available for LaTeX via CTAN as "classicthesis".

This thesis was written in accordance with recommendations of the Publication Manual of the American Psychological Association, Sixth Edition (American Psychological Association, 2010). Bibliographies were typeset by using the APA style for the biblatex-package, which is freely available via CTAN.

Figure 1 on page 22 and Figure 5 on page 79 were produced by using the LaTeX package "Ti*k*Z", which is freely available via CTAN. More Ti*k*Z examples can be found on http://www.texample.net/tikz/examples/. All other figures were produced by using ggplot2 (Wickham, 2009), which is an add-on package to R (R Development Core Team, 2011) and freely available via CRAN.

DECLARATION

Hiermit erkläre ich, dass
die Dissertation von mir selbst ohne unerlaubte Beihilfe verfasst worden ist.

Zurich, January 2012

Erich Studerus